DOCTOR OF PHILOSOPHY

The dual-targeted membrane assembly of the Streptomyces coelicolor Rieske protein

Tooke, Fiona Joy

Award date:
2016

The dual-targeted membrane assembly of the *Streptomyces coelicolor* Rieske protein

by

Fiona Joy Tooke

September 2016

Thesis submitted to the University of Dundee in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I declare that I am the author of this thesis and that, unless otherwise stated, all references cited have been consulted; that the work of which this thesis is a record of has been performed by me, and that it has not been previously accepted for a higher degree: where the thesis is based upon joint research, the nature and extent of my individual contribution is defined.

Fiona Joy Tooke

Abstract

There are two major protein translocases for protein export across the bacterial cytoplasmic membrane - the general secretory (Sec) system and the twin-arginine translocation (Tat) system. Both translocases can also insert membrane proteins. The Sec machinery inserts multiple transmembrane domains into the cytoplasmic membrane by a co-translational mechanism, whereas, the Tat machinery translocates fully folded proteins across the membrane and is only known to integrate proteins that carry a single transmembrane helix.

The Rieske protein is a membrane-bound iron-sulphur protein found in electron transport chains. In most bacteria, assembly of Rieske is Tat-dependent, as iron-sulphur cofactor insertion occurs in the cytoplasm, and the protein usually contains a single transmembrane helix anchor. However, in Actinobacteria, this protein comprises three transmembrane domains prior to the cofactor-containing domain, causing a biosynthetic insertion problem. Previous work has shown that the *Streptomyces coelicolor* Rieske protein is dual-targeted to the cytoplasmic membrane since it requires both the Sec and Tat export pathways for correct insertion.

The aim of this study was to investigate this dual-targeting mechanism and to understand the specific features of Rieske that facilitate its release by the Sec machinery and its subsequent recognition by the Tat machinery for complete insertion into the membrane. Development of reporter assays allowed assessment of the interaction of the transmembrane domains of Rieske with either the Sec or the Tat pathway. It was discovered that topological determinants, like hydrophobicity and the positive inside rule, are crucial to the Sec machinery releasing the third transmembrane domain. In comparison, the Tat machinery requires at least seven residues in the cytoplasmic loop preceding the Tat motif for efficient recognition of the membrane-tethered substrate. This work has elucidated the mechanism by which assembly of the Actinobacterial Rieske protein is co-ordinated.

Table of Contents

Table of Figures

Table of Tables

Abbreviations

Å	ångström (10^{-10} m; 0.1 nm)
aa	amino acids
Amp	ampicillin
APH	amphipathic helix
ATP	adenosine triphosphate
bp	base pair(s)
Cml	chloramphenicol
C-tail	C-terminus tail
C-terminal	carboxy terminal
C-terminus	carboxyl terminus
Da	dalton
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
EDTA	ethylenediamine tetraacetate
ER	endoplasmic reticulum
FeS	iron-sulfur
Ffh	forty-four homologue
x *g*	relative centrifugal force in multiples of standard gravity
g	gram
GTP	guanosine triphosphate
HRP	horse radish peroxidase
IM	inner membrane
Kan	kanamycin
kb	kilobase pairs (1000 bp)
kDa	kilo Daltons
Kdo	3-deoxy-D-*manno*-octulosonic acid
l	litre
LB	Luria and Bertani (medium)
LPS	lipopolysaccharide
µ	micro
M	molar
m	milli
M.I.C	minimum inhibitory concentration
mal-PEG	methoxypolyethylene glycol maleimide
MBP	Maltose binding protein

min	minute
nm	nanometre
N-terminal	amino terminal
N-terminus	amino terminus
OM	outer membrane
ΔpH	proton gradient
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMF	protonmotive force
RNC	ribosome nascent chain
rpm	rotations per minute
SAXS	small-angle X-ray scattering
s.d.	standard deviation
SDS	sodium dodecyl sulphate
Sec	secretory
SRP	signal recognition particle
Tat	twin-arginine translocase
TEMED	*N,N,N′,N′*-tetramethylethylenediamine
TF	trigger factor
TMD	transmembrane domain
Tris	tris(hydroxymethyl) aminomethane
UQ	ubiquinone
v/v	volume per volume
w/v	weight per volume

Amino Acid Abbreviations

Amino acid	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glysine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Any amino acid	Xaa	X

Acknowledgements

Firstly I would like to thank my wonderful supervisor Tracy Palmer for the opportunity to work on this project. I have really appreciated the support and encouragement over the last 4 years, especially during the frustrating times of negative data! You are absolutely inspirational and it has been incredible working for you.

Next, I would like to thank the entire Molecular Microbiology division, especially past and present TPFS members, for everyday help in the lab and creating a wonderful place to work. I am not sure I will ever work anywhere so friendly, helpful and sociable again! To those who have shared an office with me over the years, thank you for the company and conversation, science and not, which has been immensely helpful with keeping me sane. Special thanks to Grant for all your help and for the coffee and lunch entertainment. Also to Marion for whose help, support, patience and advice I am exceptional grateful. Of course, thanks to our lab manager Jackie, we would be totally lost without you.

I would like to thank those who have funded me over the years. MRC for funding my PhD studentship and Biochemical Society and Microbiology Society for proving the grants enabling me to go to some amazing conferences and workshops to present my work. Thanks to Dr Govind Chandra (JIC, Norwich) for the bioinformatics analysis.

To the friends I have made over the years, you know who you are, thank you for all the wonderful memories I will treasure forever. Special thanks to my gym buddies and coffee companions for taking breaks with me and Claudi, Adriana, Guille and Sofia for all the adventures in Scotland and abroad.

Most of all thanks to my parents and my brother, for the constant support, love and encouragement and for asking about work, even if you have no idea what I was doing. I could not have done this without you!

1 Introduction

1.1 Gram-negative bacteria

Nearly all bacteria are classified into two types; Gram-positive and Gram-negative. This classification is from a staining procedure which was developed in 1884 where Gram-positive bacteria retain the staining and Gram-negative do not. This is because of the differences in their cell envelopes.

Gram-negative bacteria, such as *Escherichia coli*, have three principle layers in their cell envelope; the outer membrane (OM), the peptidoglycan layer in the periplasm and the inner membrane (IM), as seen in Figure 1.1. The OM is a distinguishing feature of Gram-negative bacteria as it is not present in Gram-positive bacteria. The role of the OM is to protect the Gram-negative organisms from the environment, it does this by excluding toxins and providing a stabilising layer around the cell. The OM is asymmetric and composed of a phospholipid inner leaflet and a glycolipid outer leaflet, mostly composed of lipopolysaccharides (LPS) (Kamio & Nikaido, 1976). The base unit of the LPS is Lipid A, this is followed by inner and outer core polysaccharides and the O-specific polysaccharide (or O-antigen).

Lipid A is also known as endotoxin and elicits a strong immune reaction in mammals. It anchors the LPS into the OM and serves as a scaffold for assembly of the LPS. The core of the LPS can be divided into two regions: the inner core which is covalently bound to Lipid A and the outer core which is the attachment site for the O-antigen. All inner cores possess an eight-carbon sugar deoxy-D-*manno*-octulosonic acid (Kdo) which is the link to Lipid A. Together the Kdo moieties and some of the phosphate groups on Lipid A are the binding sites for divalent cations that stabilise the OM. The outer cores have a less conserved chemical structure reflecting their exposure to selective pressure, for example, *E. coli* has five known core types, R1-4 and K-12 (Raetz & Whitfield, 2002). The outer core includes heptoses, hexoses and phosphate groups which can also modify Lipid A. The O-antigen is attached to the outer core and is comprised of repetitive units, mostly monosaccharides which vary between bacterial species and strains. O-antigen

protects the cell from antibiotics which makes it an important target of the complement pathway of the immune system (Raetz & Whitfield, 2002).

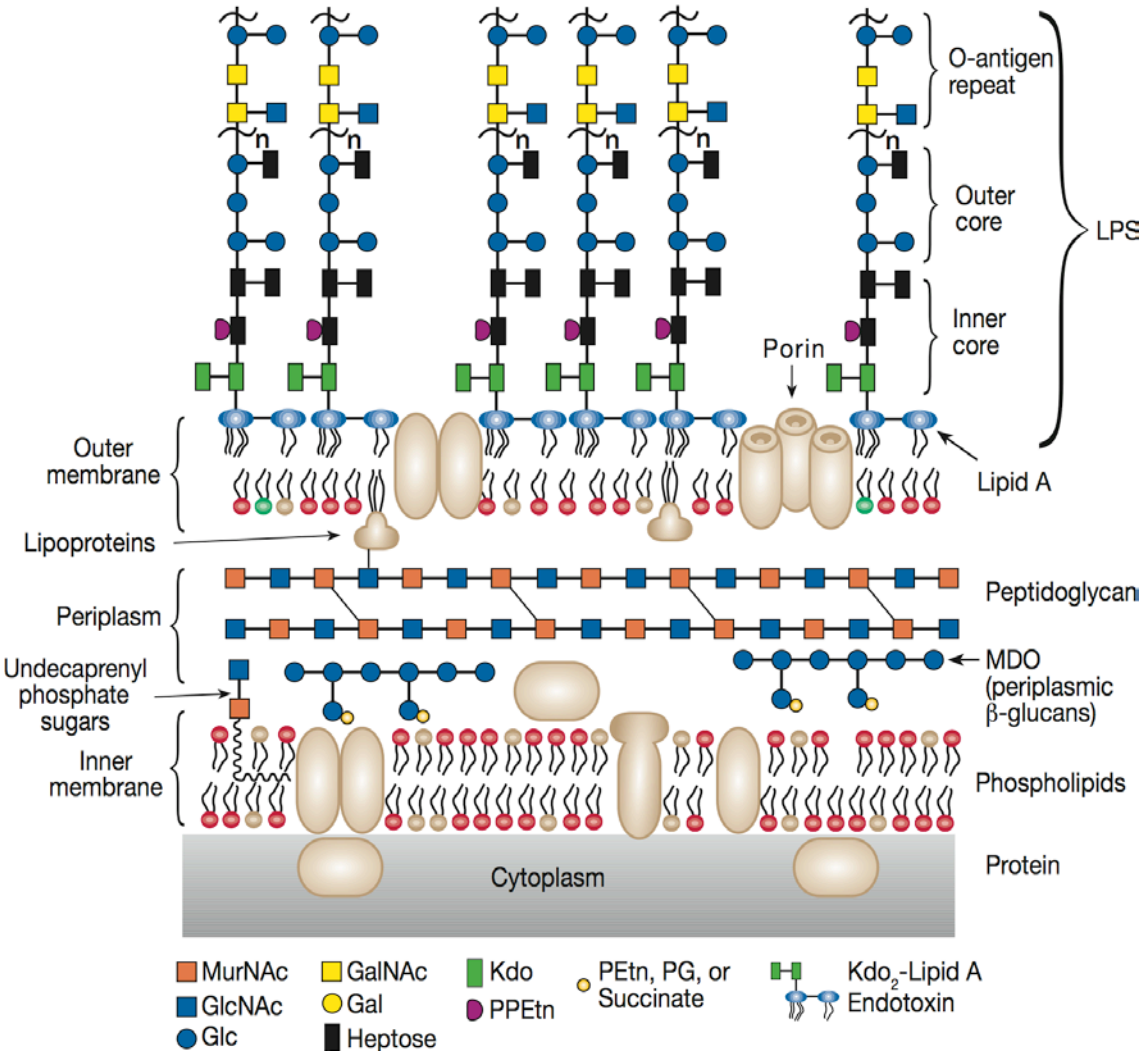


Figure 1.1 Schematic representation of the Gram-negative cell envelope.
The cell envelope of Gram-negative bacteria is composed of the cytoplasmic (inner) membrane, the periplasm and the OM. In the periplasm there is a layer of peptidoglycan and the OM of the LPS contains carbohydrate residues, coloured circles and squares. The membranes are composed of different lipids; phosphatidylethanolamine (red lipids), phosphatidylglycerol (yellow lipids), and cardiolipin (green lipids). Kdo (deoxy-D-*manno*-octulosonic acid); heptose (l-glycero-d-manno-heptose); n (variable number of O-antigen repeats); PPEtn (pyrophosphoethanolamine). Taken from Esko *et al.* (2009).

Proteins in the OM can generally be divided into two classes, lipoproteins and β-barrel proteins. Lipoproteins contain lipid moieties which are thought to embed the protein in the inner leaflet of the OM, therefore, lipoproteins are not believed to be transmembrane proteins. Most transmembrane OM proteins classically adopt β-barrel transmembrane structures to allow the passage of molecules. A large number of these are porins which are integral pore-forming proteins that function to mediate the transport of small nutrients and ions across the membrane with differing selectivity (Nikaido, 2003).

In comparison to the OM, the IM, also known as the cytoplasmic membrane, consists of a symmetric bilayer of phospholipids. *E. coli* membranes contain approximately 75% phosphatidylethanolamine (PE), 20% phosphatidylglycerol (PG), and 5% cardiolipin (CL) (Cronan, 2003). This bilayer is semi-permeable and serves as an electrochemical barrier. Although some small molecules like O_2, CO_2 and H_2O can freely diffuse through it, other molecules are transported by multiprotein complexes (Silhavy *et al.*, 2010). Proteins present in the IM can be either peripherally associated with the surface or span the bilayer. Many proteins involved in energy production, such as cytochromes and ATP synthases, are localised in the IM, therefore, this is where all the energy production for the cell takes place during respiration. Translocation machineries are also present to enable proteins to be integrated into the membrane or transported into the periplasm.

The periplasm is the aqueous space located between the IM and the OM and it contains the peptidoglycan cell wall. Peptidoglycan consists of parallel strands of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) connected by β-1,4 glycosidic bonds (Vollmer *et al.*, 2008). The rigidity of the peptidoglycan determines cell shape. The peptidoglycan wall is only a few layers thick in *E. coli* and is covalently linked to lipoproteins in the OM. The periplasm is more viscous than the cytoplasm and is densely packed with proteins (Mullineaux *et al.*, 2006). Some of these proteins are involved in cell wall biogenesis and division, nutrient uptake, motor protein assembly and anaerobic respiration.

1.1.1 *E. coli* as a model organism

E. coli is a Gram-negative, rod-shaped bacteria belonging to the family Enterobacteriaceae in the phylum of γ-proteobacteria. The *E. coli* bacterium is typically found in the mucus layer of the mammalian colon and it is the most abundant facultative anaerobe of the human intestinal microflora. Although most naturally occurring strains are harmless, there are some pathogenic strains which cause diarrheal diseases, urinary tract infections and meningitis, for example, enteropathogenic *E. coli* (EPEC) and enterhaemorrhagic *E. coli* (EHEC). Most virulent strains arise through the acquisition of virulence genes, such as those encoding the Shiga toxin, and can cause illness in even the healthiest host (Johnson & Nolan, 2009).

The genetic tractability, fast reproduction time and simple growth requirements has enabled *E. coli* to be used as model organism and it has become one of the best characterised prokaryotes. The full sequencing of its entire genome has been reported, contributing to a better understanding of its genetic and phenotypic diversity. The size of the genome varies between strains; standard lab strains have genomes of ~4.5 million base pairs and 4000 genes compared to pathogenic strains with over 5.9 million base pairs and 5500 genes (Blattner *et al.*, 1997, de Muinck *et al.*, 2013, Lukjancenko *et al.*, 2010). *E. coli* K-12 has become a model bacterium used for biochemical and cellular studies.

1.2 Gram-positive bacteria

The Gram-positive cell envelope differs from the Gram-negative in several key ways. Primarily there is no OM to protect the organism. Instead, the Gram-positive bacterial IM (cytoplasmic membrane) is surrounded by layers of peptidoglycan, much thicker than is found in Gram-negative bacteria to withstand the high intracellular pressure. Threading through the layers of peptidoglycan are long anionic polymers known as teichoic acids, which are composed of glycerol phosphate, glycol phosphate or ribitol phosphate repeats. A schematic representation of the Gram-positive cell envelope is shown in Figure 1.2.

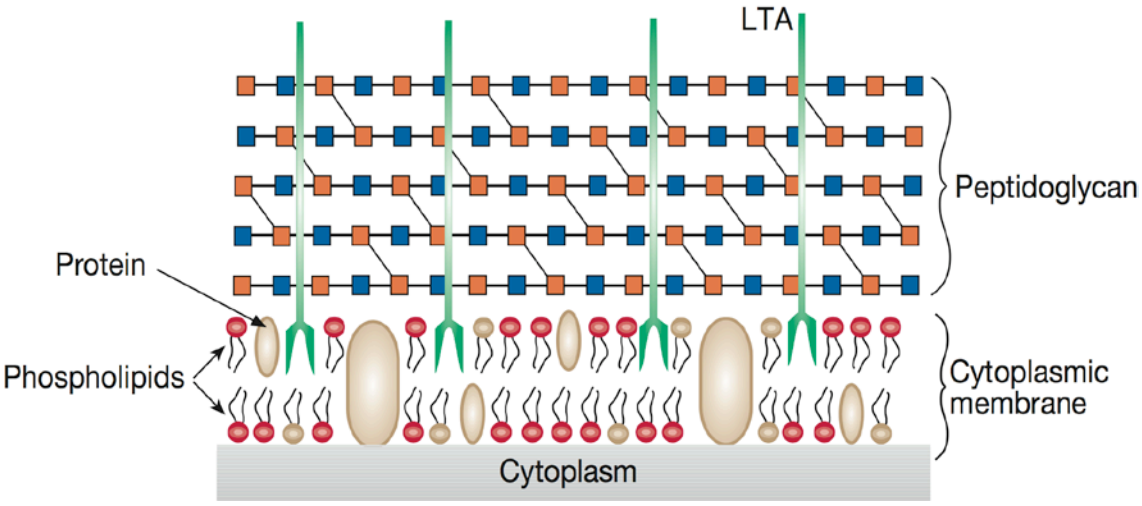


Figure 1.2 Schematic representation of the Gram-positive cell envelope.
The cell envelope of Gram-positive bacteria is composed of the cytoplasmic (inner) membrane and a peptidoglycan layer. The membrane is composed of different lipids; phosphatidylethanolamine (red lipids), phosphatidylglycerol (yellow lipids) and LTA (lipoteichoic acids, in green). Taken from Esko *et al.* (2009).

As in Gram-negative bacteria, the peptidoglycan is composed of disaccharide-peptide repeats coupled through glycosidic bonds to form linear glycan strands crosslinked into a mesh-like framework (Silhavy *et al.*, 2010). However, unlike Gram-negative cells, the Gram-positive peptidoglycan contains many layers and is 30-100nm thick (10-20 layers). The specific crosslinks between the glycan strands can differ among organisms (Vollmer *et al.*, 2008).

Teichoic acids are anionic cell surface polymers, there are two different types; wall teichoic acids (WTA) and lipoteichoic acids (LTA). WTAs are coupled to peptidoglycan and extend perpendicular to the peptidoglycan mesh, whereas LTAs are anchored to the cell membrane. Structural variations in WTAs are extensive but the most common types are composed of a disaccharide linkage unit joined to a polyribitol phosphate or polyglycerol phosphate chain. LTAs are similar to WTAs but contain polyglycerol phosphate repeats of opposite chirality. They are anchored to membrane-embedded glycolipids and so extend from the cell surface into the peptidoglycan layers. Both teichoic acids comprise a continuum of ionic charge which starts at the cell surface but extends beyond the peptidoglycan layer. Although neither WTAs nor LTAs are essential, deletion of both pathways is not possible as the genes involved are synthetic lethal (Oku *et al.*, 2009, Morath *et al.*, 2005). The functions of teichoic acids are varied and species-dependent. The anionic nature allows them to bind cations and play a role in cation homeostasis (Marquis *et al.*, 1976). The networks of metal cations between WTAs can influence the rigidity of the cell wall and may be involved in scaffolding or activating hydrolytic enzymes. Collectively these polymers can contribute to over 60% of the mass of the Gram-positive cell wall, indicating their importance to structure and function.

In addition to teichoic acids, there are a variety of surface proteins some of which are analogous to those within the Gram-negative periplasm (Dramsi *et al.*, 2008). All of these proteins are retained in or near the membrane. This can be through membrane-spanning helices, lipid anchors in the membrane, covalently attached to the peptidoglycan or bound to teichoic acids (Scott & Barnett, 2006). These proteins are key to many functions of the cell, including; adhesion, invasion, evasion and virulence.

Within the phylum of *Actinobacteria*, there is a suborder of *Corynebacterineae* which includes the genera *Corynebacterium*, *Mycobacterium* and *Nocardia*. These bacteria are generally classed as high G+C Gram-positive bacteria, however, their cell envelope also has characteristics of Gram-negative bacteria. A genome-based phylogeny places them

between Gram-positive and Gram-negative (Fu & Fu-Liu, 2002). The cell envelope of these bacteria is very complex and this contributes to their virulence.

The peptidoglycan layer contains covalently attached arabinogalactan which is covalently attached to mycolic acids. Mycolic acids have long alkyl side chains giving a waxy appearance, which along with the high density of lipids present, interferes with the Gram staining test. Corynebacterineae have an OM which seems to be symmetrical, unlike the Gram-negative OM, containing the essential mycolic acids and potentially similar lipids in both leaflets (Hoffmann *et al.*, 2008). This can be surrounded by a non-covalently linked outer capsule of proteins and polysaccharides.

M. tuberculosis is the aetiological agent of tuberculosis and a member of the Corynabacterineae suborder. The unique bacterial cell envelope is key to its success as a pathogen. When in a host the envelope can be remodelled, this allows the generation of a population of cells which differ in composition of their cell envelopes, potentially increasing their survival *in vivo* (Seiler *et al.*, 2003, Jain *et al.*, 2007). The extractable lipids of the cell envelope can act as direct effectors of pathogenesis by modulating the host immune responses or altering intracellular trafficking. These differences to other bacteria could be useful as targets for inhibitors to treat human disease.

1.3 Actinobacteria

The phylum *Actinobacteria* is comprised of bacteria with a relatively high Guanine and Cytosine content (>55 mol% in genomic DNA) and so are known as high G+C Gram-positive bacteria. They are one of the largest phyla within Bacteria and exhibit vast diversity in terms of their morphology, physiology and metabolic capabilities. Actinobacteria are ubiquitously distributed in both aquatic and terrestrial ecosystems, nonetheless, they are most abundant as saprophytic, soil-dwelling organisms that spend most of their life cycle as semidormant spores, especially if nutrients are limited. Soil populations are dominated by the genus *Streptomyces* accounting for over 95% of Actinomycetales strains isolated from soil. Actinobacteria prefer alkaline soil and soils rich in organic matter and are found both on the soil surface and at depths of more than 2 m below ground (Goodfellow & Williams, 1983). A number of Actinobacteria live in close association with higher organisms as a component of different microbiomes, contributing to more than a third of healthy human microbiota.

The life cycle of Actinobacteria is complex, most have a mycelial lifestyle and usually reproduce by forming asexual spores, as seen in Figure 1.3. Actinobacteria form a substrate mycelium but on solid surfaces may differentiate to aerial hyphae which produces reproductive spores. The morphologies of Actinobacteria differ with respect to the presence or absence of a substrate mycelium or aerial mycelium, the colour of the mycelium, the production of diffusible melanoid pigments, and the structure and appearance of their spores (Barka *et al.*, 2016).

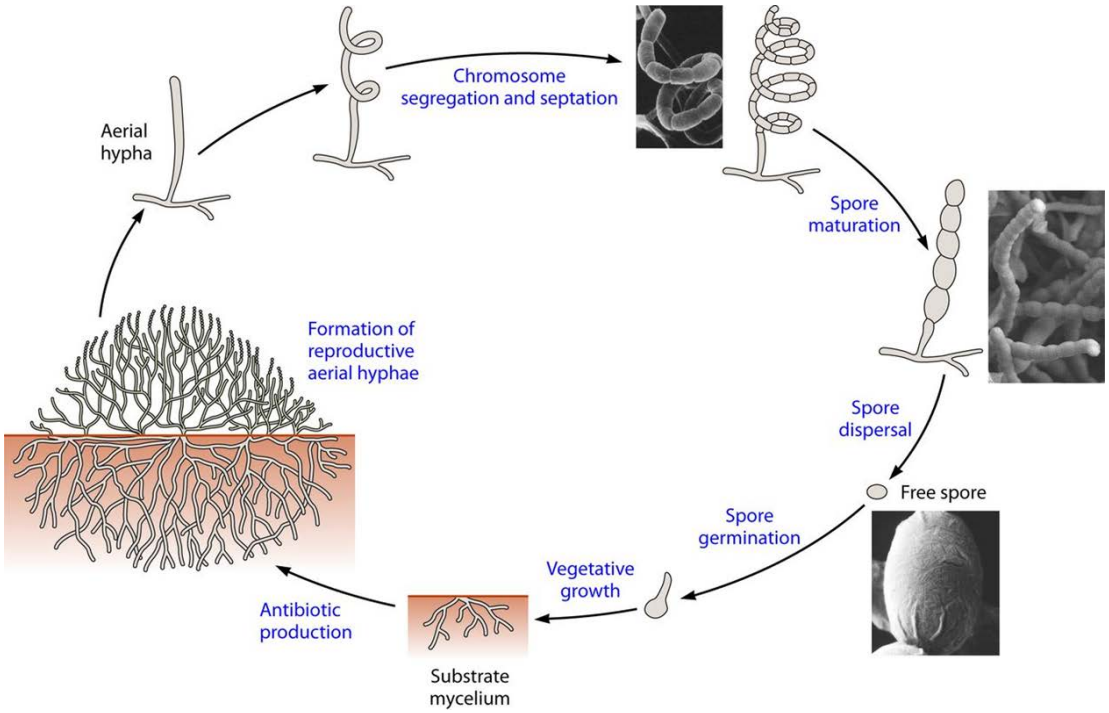


Figure 1.3 Schematic representation of the life cycle of sporulating Actinomycetes.
The life cycle begins with spore germination and outgrowth of the substrate mycelium. Reproductive aerial hyphae are produced in response to nutritional and other stress signals. These then undergo cell division to form spores. Taken from Barka *et al.* (2016).

Actinobacteria are of high importance in the field of biotechnology as they produce an abundance of bioactive secondary metabolites which have extensive industrial, medical and agricultural importance. In particular, Actinomycetes produce approximately two-thirds of all known antibiotics, the majority of which are produced by *Streptomyces*. There are also pathogenic strains which cause severe human and animal diseases as well as agricultural losses, for example; *M. tuberculosis, Corynebacterium diphtheriae*, and *Streptomyces scabies*.

1.3.1 *Streptomyces*

The genus *Streptomyces* includes over 150 known species, they are abundant and important in the soil and play major roles in the cycling of carbon trapped in insoluble organic debris by secreting a variety of hydrolytic exoenzymes. *Streptomyces* also produce a spectacular multitude and diversity of bioactive secondary metabolites which are used to compete with other microorganisms. Many of these metabolites have

beneficial properties and consequently are of great interest in medicine and industry, for example, antibiotic compounds.

Streptomyces coelicolor is a model system for studying Actinobacteria and is known to produce a plethora of different antibiotic compounds, currently around 10000 bioactive products. When the genome sequence was published in 2002 for *S. coelicolor* A3(2) (Bentley *et al.*, 2002) over 20 previously-undiscovered biosynthetic gene clusters for secondary metabolites were identified (Challis & Hopwood, 2003). This genome also contains an unprecedented proportion of regulatory genes, most likely to be involved in responses to external stimuli and stresses, possibly explaining why not all metabolites have been discovered previously. This indicates that there is potential for new antibiotics, anti-tumour drugs and other beneficial molecules from these organisms.

Genome analysis has revealed a synteny between the central 'core' of the *S. coelicolor* chromosome and the whole chromosome of pathogens *M. tuberculosis* (Cole *et al.*, 1998) and *C. diphtheriae* (Cerdeno-Tarraga et al., 2003) indicating a potential for extrapolating data from *S. coelicolor* into other Actinobacteria.

1.3.2 *Mycobacterium*

The genus *Mycobacterium* harbours some of the most notorious human pathogens including *M. tuberculosis* and *M. leprae*, the agents responsible for tuberculosis (TB) and leprosy, respectively. Mycobacterial diseases are often associated with immunocompromised patients especially those with HIV. The pathogenic success of these bacteria is in their unusual cell envelope which is essential for intracellular survival and enables Mycobacteria to resist a number of antibiotics.

TB is spread through the air, and *M. tuberculosis* can remain airborne in water droplets for hours. It only requires two or three bacteria to trigger the innate immune response where either the host will kill the bacterium, immediately develop an infection or the bacterium will remain latent within the body. *M. tuberculosis* can survive and multiply within macrophages which can lead to pulmonary disease or hematogenous

dissemination and potential seeding of multiple organs, which may eventually give rise to extrapulmonary disease (Getahun *et al.*, 2015).

The World Health Organisation (WHO) has predicted that nearly one third of the world has an asymptomatic or latent TB infection and TB now ranks alongside HIV as a leading cause of death worldwide. The Global tuberculosis report for 2015 (http://www.who.int/tb/publications/global_report/en/) states that in 2014, TB killed 1.5 million people (1.1 million HIV-negative and 0.4 million HIV-positive). Effective drug treatments for TB is a six-month regimen of four first-line drugs: isoniazid, rifampicin, ethambutol and pyrazinamide. *M. tuberculosis* bacteria are becoming resistant to some of these drugs, isoniazid and rifampicin, with 480000 cases of multidrug-resistant TB (MDR-TB) estimated to have occurred in 2014. New TB drugs are in development and TB vaccines are also being investigated.

1.4 Protein Transport

Different proteins are located to specific compartments of the bacterial cell. However, all of these proteins are synthesised by ribosomes in the cytoplasm and cannot cross through or insert into membranes by themselves. Therefore, there are specific protein transport machineries which translocate proteins into their designated compartment.

Protein secretion is the movement of proteins from the cytoplasm through the IM and OM to the extracellular medium or into another target cell. This process is carried out by specialised Type I to Type VI secretion systems. In contrast, protein translocation is the movement of proteins across the IM by either the general secretory (Sec) pathway, the YidC-dependent integration pathway or the twin-arginine translocase (Tat) pathway.

1.5 The general secretory (Sec) pathway

In bacteria, many proteins undergo transport across the cytoplasmic membrane into the periplasm using the general secretory (Sec) pathway. This is the main route that proteins take to reach the periplasm and is conserved across all domains of life (Pohlschroder *et al.*, 1997). The components of the *E. coli* Sec pathway were identified during the 1980s by genetic screening. A number of mutants were isolated exhibiting defective protein secretion (*sec*) and dominant suppressor mutant alleles of protein localisation (*prl*) genes which restored the export of substrates with altered signal peptides (Bieker *et al.*, 1990, Trun *et al.*, 1988, Danese & Silhavy, 1998). These mutants led to the discovery of two Sec protein complexes involved in the export of proteins (Bieker-Brady & Silhavy, 1992) which facilitated the reconstitution of the Sec translocase *in vitro* (Brundage *et al.*, 1990, Akimaru *et al.*, 1991, Bassilana & Wickner, 1993).

1.5.1 Sec signal peptides

Proteins destined for export by the Sec pathway are synthesized as precursor proteins with an N-terminal signal peptide that targets them to the translocase. These signal peptides do not share sequence homology but have clearly definable features, as seen in Figure 1.4. They are usually 18-26 residues in length with a conserved tripartite structure. The n-region contains at least one positive charge and is followed by a core hydrophobic (h-) region and a polar c-region. These features are important for targeting to the Sec pathway; if the positive charges in the n-region are removed then Sec translocation is inhibited, in comparison, increasing the number of positive charges increases the efficiency of translocation (Sasaki *et al.*, 1990). The c-region contains a signal peptidase cleavage site, usually an AxA motif, where A is alanine and x is any amino acid. Mutations to this site can inhibit cell growth and protein processing, due to the absence of mature Sec-dependent proteins (Barkocy-Gallagher & Bassford, 1992). The hydrophobicity of the h-region is crucial for efficient translocation and disruption of this hydrophobic core leads to defects in translocation (von Heijne, 1990). This region is also involved in targeting to the different Sec translocation routes; post-translationally exported proteins recognised by the molecular chaperone SecB or co-translationally exported proteins recognised at the ribosome by the signal recognition particle (SRP). Signal peptides with a more hydrophobic h-region target to the SRP-dependent pathway, whereas less hydrophobic h-regions will target to the SecA/SecB-dependent route for export through the Sec machinery (De Gier *et al.*, 1997).

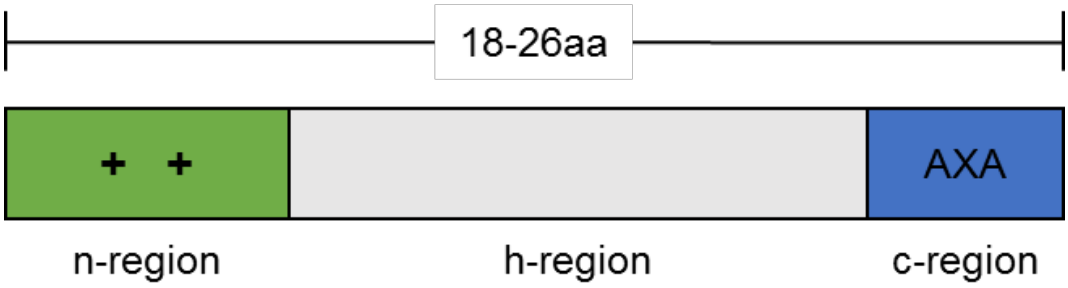


Figure 1.4 General structure of Sec signal peptides.
The N-terminal Sec signal peptide has a tripartite structure; a positive n-region (green), a hydrophobic h-region (grey) and a polar c-region (blue). The signal peptidase AxA cleavage site is located within the c-region.

1.5.2 Components of the Sec translocon

The *E. coli* Sec translocon is heterotrimeric and consists of two universally conserved proteins SecY and SecE. A third protein SecG was identified by co-purification with the SecYE complex and is not essential for translocation or cell viability but has been seen to increase the efficiency of translocation, especially in low temperatures (Nishiyama *et al.*, 1993, Nishiyama *et al.*, 1994, Hanada *et al.*, 1994). In eukaryotes the Sec machinery consists of Sec61αβγ and in archaea SecYEβ. The first high resolution insight into the Sec translocon came from the crystal structure of the archaeon *Methanococcus jannaschii* SecYEβ complex presenting the complex in a closed conformation, seen in Figure 1.5 (Van den Berg *et al.*, 2004).

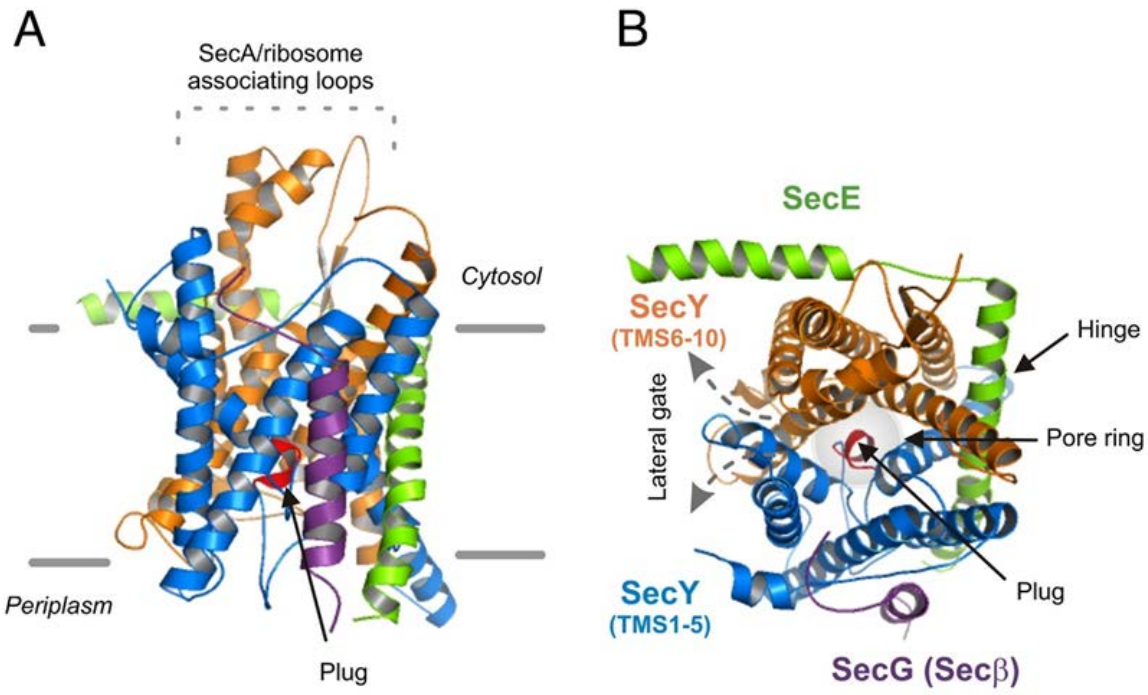


Figure 1.5 Structure of the *M. jannaschii* Sec translocase.
A) Side view and B) top view (cytosolic face) of the crystal structure obtained from SecYEβ of *M. jannaschii* (Van den Berg *et al.*, 2004). The two halves of SecY are indicated in blue and orange with the plug domain in red. SecE is in green and SecG (Secβ) in purple. Figure taken from Du Plessis *et al*. (2010).

From the structure (Figure 1.5) it can been seen that SecY forms the translocation channel, consisting of 10 transmembrane domains (TMDs) which are divided into two halves (TMDs 1-5 and 6-10) that are pseudo-symmetrically aligned to generate a clamshell-like structure. An external loop acts as a hinge and connects TMD5 and TMD6 at the back of the molecule. The TMDs form an hourglass shape which is constricted, approximately halfway across the membrane, by six hydrophobic residues. This is a seal that prevents the leakage of ions and may form a hydrophobic ring around the translocating polypeptide (Park & Rapoport, 2011, Gumbart & Schulten, 2008, Saparov *et al.*, 2007). Structural data has shown that the protruding loops in the cytoplasmic face, between TMD6-7 and TMD8-9, are involved in the interactions of SecY with its partner protein SecA and the ribosome, indicated in Figure 1.5A (Frauenfeld *et al.*, 2011, Menetret *et al.*, 2007, Zimmer *et al.*, 2008).

A plug domain (Figure 1.5, in red), located on the periplasmic side, is formed by an extension of TMD2a and prevents the passage of ions. Deletion of this plug generates fluctuations in the channel between closed and open conformations indicating its importance in stability (Saparov *et al.*, 2007). The plug is maintained in the closed state, however, during transport it undergoes a conformational change and is displaced. To allow the insertion of proteins into the IM, the SecY translocon opens laterally between TMD2-3 and TMD7-8. This conformational change occurs during protein translocation when the signal peptide interacts with TMD2b and TMD7 (Plath *et al.*, 1998, du Plessis *et al.*, 2011).

In *E. coli*, SecE is an essential 3TMD protein that enwraps the SecY channel by contacting both sides of the lateral gate. TMD1 and TMD2 are not essential and are not present in most eubacterial SecE proteins, however, the conserved cytoplasmic region is found to be essential for function (Murphy & Beckwith, 1994). When SecE is underexpressed compared to SecY or when the latter is overexpressed, SecY is degraded by FtsH which can lead to an effect on membrane integrity, this indicates that SecE stabilises the SecY channel (Kihara *et al.*, 1995).

1.5.3 Post-translational Sec translocation

Proteins destined for the periplasm or the OM undergo post-translational targeting and translocation using the SecA/SecB pathway. This decision occurs at the ribosome exit tunnel where a cytosolic chaperone, Trigger Factor (TF), is recruited. TF interacts with the ribosome subunits L23 and L29 which are located at the ribosome exit tunnel and with the nascent chain maintaining it in an unfolded conformation (Kramer *et al.*, 2002, Valent *et al.*, 1995). The recognition of the signal peptide of the nascent polypeptide by TF recruits the SecB chaperone which binds the precursor protein in its mature region (Gannon *et al.*, 1989). Studies of the precursor form of maltose binding protein (MBP) have found that in the absence of SecB, MBP will fold into a stable tertiary structure which is unable to translocate through the Sec translocon, consequently, the binding of SecB prevents the folding of the precursor protein before translocation (Bechtluft *et al.*, 2007, Collier *et al.*, 1988). The dual functionality of SecB also involves a specific interaction with SecA which targets the precursor to the Sec translocon. SecB has been crystallised from *Haemophilus influenzae* and *E. coli* both structures showing that SecB is organised as a dimer of dimers, with two peptide binding grooves on each side (Xu *et al.*, 2000, Dekker *et al.*, 2003).

The motor component of the Sec pathway is SecA, which recognises SecB *via* 22 conserved residues in its extreme C-terminus (Fekkes *et al.*, 1997) and also binds to the precursor through both its mature region and its signal peptide. SecA interacts with the SecYEG complex guiding both the precursor and SecB to the translocon (Hartl *et al.*, 1990). The precursor is then transferred from SecB to SecA in an energy-independent process and SecB is released upon the binding of ATP to SecA, it is this step which initiates translocation (Fekkes *et al.*, 1997).

SecA is a 204kDa ATPase which can be found in both soluble and membrane-bound forms. In addition to the C-terminal tail and zinc finger regions that are involved in the recognition of SecB, SecA has an N-terminal DEAD helicase motor domain which is essential for protein translocation (as the site of ATP binding and hydrolysis), and a

preprotein binding domain (Sianidis *et al.*, 2001, Papanikou *et al.*, 2005). The translocation of the precursor through the Sec translocon is driven by ATP hydrolysis and enhanced by the PMF.

Translocation of a protein through the Sec pathway has three main steps; sorting and targeting (which is discussed above), translocase priming and activation, and the translocation of the precursor. Chatzi *et al.* suggested a model of SecA-mediated post-translational translocation through the Sec pathway (Chatzi *et al.*, 2014). In this model, ADP-SecA is presented as a tight dimer in the cytoplasm and one subunit proceeds to dock onto the SecYEG complex. The preprotein, with or without the chaperone SecB, interacts with the SecA-SecYEG complex. Tight signal peptide binding to SecA promotes triggering where the activation energy state of the SecA-SecYEG complex is lowered, followed by an allosteric change of the conformation of the translocase (Gouridis *et al.*, 2009). The preprotein is then trapped by the SecA dimer which monomerises and catalyses ATP hydrolysis. This step, with the PMF, initiates the translocation of the signal sequence into the SecYEG translocon. The interaction of the signal sequence with the lateral gate of SecY causes the flipping of the periplasmic plug domain to allow the full translocation of the polypeptide across the membrane (Tam *et al.*, 2005). The PMF is involved in different stages of the translocation process, including the speeding up of the dissociation of SecA, either by conformational change in SecY or promoting ADP release from SecA. Cycles of ATP binding and hydrolysis alongside SecA dissociation and rebinding cause the translocation of the polypeptide across the membrane. Once the signal sequence reaches the periplasm it can be cleaved by the signal peptidase. A simplified model of this translocation process is presented in Figure 1.6.

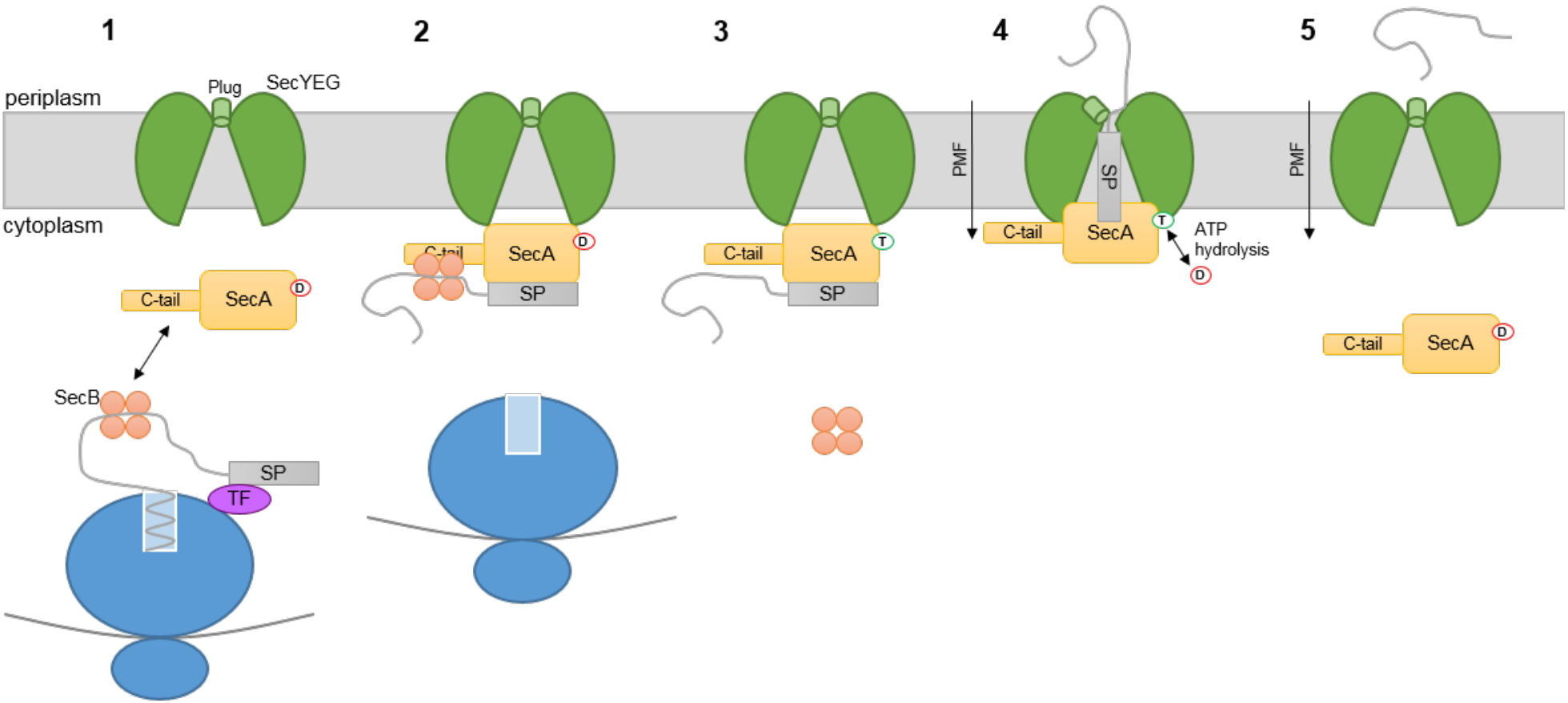


Figure 1.6 Simplified model of SecA/SecB-dependent post-translational translocation.
1) Trigger factor (TF) binds the ribosome near the tunnel exit and interacts with the signal peptide (SP) of the nascent polypeptide. The chaperone SecB is recruited, binding the mature domain of the polypeptide to prevent premature folding. 2) The C-tail of the ATPase SecA binds SecB and the signal peptide of the polypeptide and recruits both to the SecYEG translocon.3) Following docking to the SecYEG complex, ATP binding to SecA elicits the release of SecB. 4) SecA hydrolyses ATP, initiating the insertion of the polypeptide into SecYEG. The proton motive force (PMF) and the interaction of the polypeptide with the lateral gate of SecY flips the plug domain causing the channel to open. 5) After translocation the signal peptide is cleaved and the mature protein is released into the periplasm and SecA dissociates from SecYEG translocon. ADP is indicated by a D in a red circle and ATP by a T in a green circle.

1.5.4 Co-translational Sec translocation

The Sec pathway is also used to insert proteins into the IM independent of SecA/SecB. Instead the nascent polypeptide is recognised by the signal recognition particle (SRP) as it emerges from the ribosome, this complex is then targeted to the SecYEG complex for co-translational translocation to occur. A simplified scheme for the targeting of a protein with a single TMD, as a signal anchor, to the Sec complex is described in Figure 1.7.

The bacterial SRP is composed of the protein Ffh (fifty-four homologue) bound to 4.5S RNA generating a ribonucleoprotein which is essential in bacteria for translocation of proteins (Phillips & Silhavy, 1992). Ffh contains two functional domains connected by a linker. The C-terminal M-domain and an NG-domain which contains the N-terminal domain packed tightly against a central G-domain. The M-domain binds SRP RNA and the signal peptide (Hainzl *et al.*, 2011, Keenan *et al.*, 1998, Janda *et al.*, 2010). The NG-domain binds the ribosomal subunit L23 at the tunnel exit site, contains the GTPase site and interacts with the SRP receptor (Freymann *et al.*, 1997, Halic *et al.*, 2006).

SRP primarily docks to the L23 subunit of the ribosome close to the polypeptide exit tunnel where it scans the emerging nascent chains for hydrophobic sequences (Gu *et al.*, 2003). The hydrophobic TMD segment or h-region of a signal peptide binds to a deep hydrophobic groove in the M-domain (Janda *et al.*, 2010, Hainzl *et al.*, 2011). This leads to the formation of the Ribosome Nascent chain (RNC) complex containing the SRP bound to the ribosome and the emerging polypeptide. This complex is targeted to the membrane and interacts with the SRP receptor, FtsY, a bacterial peripheral membrane protein. FtsY interacts with SecY through cytoplasmic loops C4 (between TMD6-7) and C5 (between TMD8-9) which targets the RNC complex to the SecYEG translocon (Angelini *et al.*, 2005, Kuhn *et al.*, 2011).

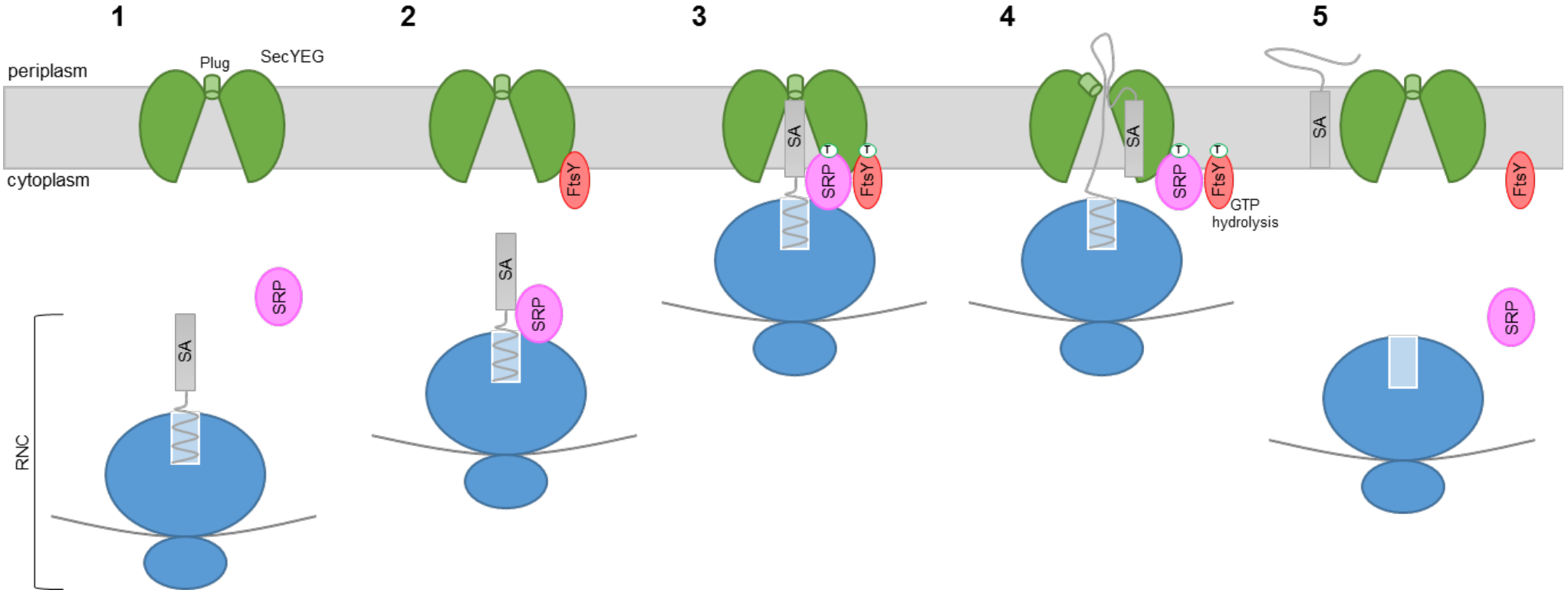


Figure 1.7 Simplified model of SRP-dependent co-translational translocation.
1) During translation a hydrophobic signal anchor (SA) emerges from the ribosome exit tunnel forming a ribosome nascent chain complex (RNC). The signal recognition particle (SRP) is recruited. 2) The SRP binds to SA of the nascent polypeptide and the complex targets to the IM. 3) SRP interacts with its receptor (FtsY) and docks the complex onto the SecYEG translocon 4) GTP hydrolysis occurs for SRP and FtsY, stimulated by their interaction. SRP is released from the complex allowing the insertion of the polypeptide chain. The SA is inserted into the membrane through the lateral gate of SecYEG. 5) Following translocation the components are released. GTP is signified by T in a green circle.

The Ffh domain forms a dimer with FtsY with the interaction of their highly homologous NG-domains (Luirink *et al.*, 1994, Egea *et al.*, 2004). This dimerization causes conformational changes between open, early, closed and activated states which stimulates reciprocal GTPase activity of both NG-domains. The GTP hydrolysis allows the release of the RNC by weakening the SRP interactions and transfers the nascent chain to SecY (Angelini *et al.*, 2005, Egea *et al.*, 2004, Shan *et al.*, 2007). The ribosome exit tunnel lines up with the SecY channel allowing translocation of the emerging polypeptide (Menetret *et al.*, 2007). Nascent polypeptide insertion triggers lateral gate opening between TMD2b and TMD7 allowing TMDs to escape into the membrane (Egea & Stroud, 2010).

Membrane proteins are targeted to the translocon as outlined in Figure 1.7. The insertional orientation of Sec-dependent proteins relies on topology determinants, Figure 1.8 shows a schematic representation of the co-translational insertion of an N_{in}-C_{out} membrane protein promoted by the positive charges in the N-terminus which remain in the cytoplasm. When a hydrophobic segment is inserted into the SecYEG translocon it interacts with the lateral gate which opens by 12-14 Å (Egea & Stroud, 2010, Gogala *et al.*, 2014). The interactions of the hydrophobic signal peptide with the mammalian Sec61 channel have recently been unlocked using cryo-EM reconstructions.

Previously it was found that in the quiescent state, the Sec channel is closed both laterally towards the lipid bilayer and axially across the membrane, as shown in Figure 1.5 (Van den Berg *et al.*, 2004). However, a structure containing the bound ribosome indicates that the interaction with the ribosome primes the channel, shown by a conformational change which causes a slight opening of the lateral gate and destabilisation of the channel (Voorhees *et al.*, 2014). A functional translocation intermediate has been reconstructed through cryo-EM to show the RNC stalled at position 86 of the preprotein. This structure shows the location of the signal peptide between TMD2 and TMD7 of Sec61α, near the lateral gate, and the displacement of the plug domain (Voorhees & Hegde, 2016).

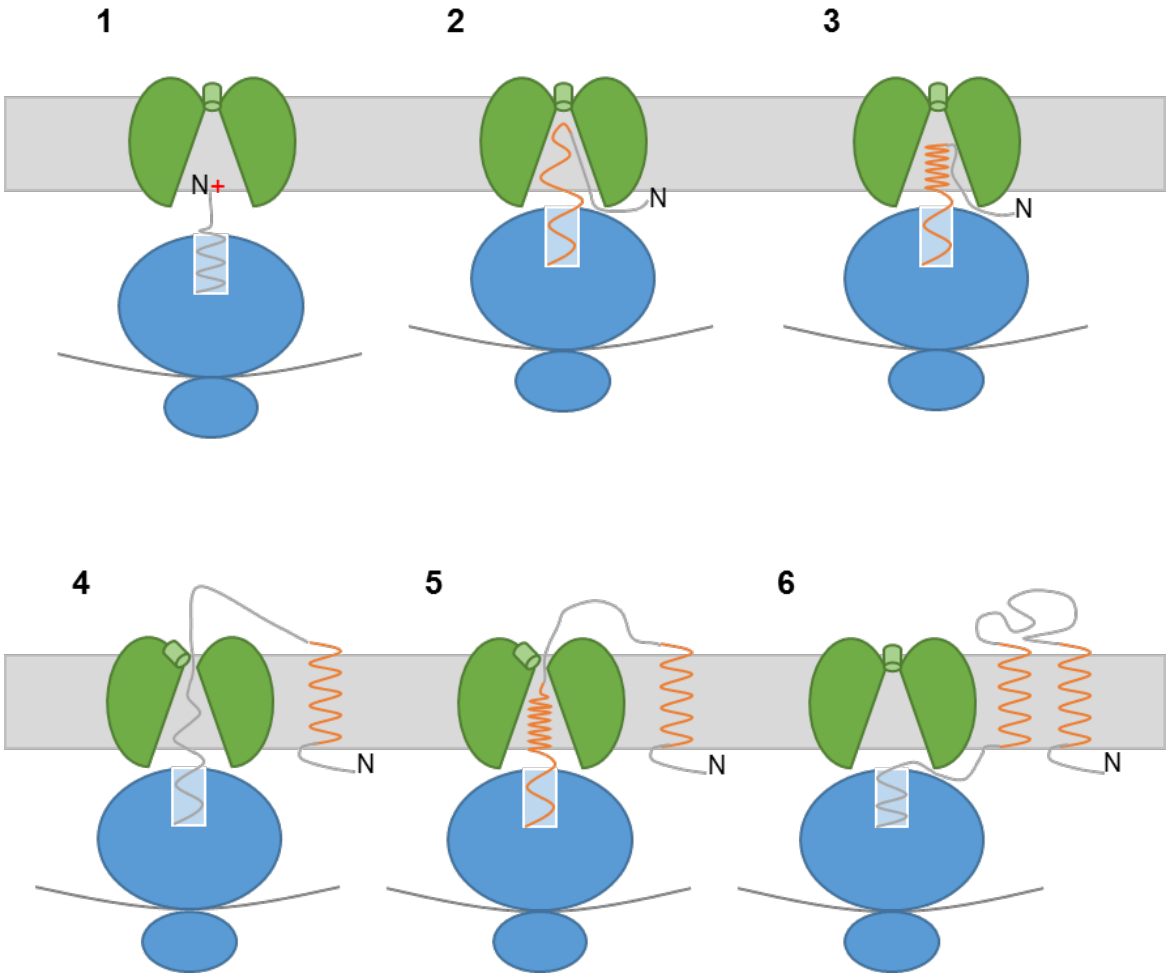


Figure 1.8 Schematic of co-translational Sec translocation.
1) The hydrophobic nascent chain enters the Sec translocon directly from the ribosome. 2) Hydrophobic residues open the lateral gate within the translocon. 3) This allows the hydrophobic α-helix to contact the lipid bilayer and escape into the membrane as a transmembrane domain. 4) The plug is displaced by hydrophilic stretches of chain allowing it to cross the membrane into the periplasm. 5) The next hydrophobic stretch starts the process again by opening the lateral gate. 6) It then escapes into the membrane and the next hydrophilic stretch is released into the cytoplasm.

From this it can be inferred that the hydrophobic segment, positioned in the region of the lateral gate, has energetically favourable interactions with the lipid bilayer, which is the driving force of intercalation into the lateral gate and the displacement of the plug domain, thus, producing a continuous translocation pore (Voorhees & Hegde, 2016). Therefore, the signal-engaged Sec61 is open both axially across the membrane and laterally towards the lipid bilayer. This open state can allow the hydrophobic segments to escape into membrane possibly through the hydrophobic effect, where the TMDs are partitioned

into the lipid bilayer using the thermodynamic partitioning model which is supported by hydrophobicity scales (Hessa *et al.*, 2005, White & von Heijne, 2008).

In comparison, hydrophilic segments are retained in the translocon and do not have such an effect on the conformation of the channel or lateral gate (Gogala *et al.*, 2014). Non-cytoplasmic loops and domains are translocated into the bacterial periplasm or the ER lumen with long cytoplasmic regions probably translocated by a mode similar to post-translation translocation, continuing until the next hydrophobic segment interacts with the lateral gate. Cytosolic domains and loops are released through a gap of 10-20 Å on one side of the ribosome-translocon junction (Frauenfeld *et al.*, 2011, Park *et al.*, 2014, Beckmann *et al.*, 2001).

1.5.5 SecDF-YajC accessory proteins

SecD, SecF and YajC are membrane components which are encoded within the *secD* operon (Gardel *et al.*, 1990). Together they from a membrane complex which associates transiently with the Sec translocon and is thought to stimulate translocation (Duong & Wickner, 1997a). Although SecD and SecF are not essential, in *E. coli*, their inactivation causes both a severe protein secretion defect and a severe growth inhibition, especially at low temperatures (Pogliano & Beckwith, 1994). Initially these proteins were implicated in the post-translational pathway and the cycling of SecA (Duong & Wickner, 1997b). They are also thought to be involved in PMF-dependent translocation, however, the absence of SecDF does not prevent this translocation (Nouwen *et al.*, 2001).

The *Thermus thermophilus* SecDF crystal structure suggests that SecDF is involved in the transfer of protons and uses the PMF to undergo conformational changes which helps to pull the translocating polypeptide from the SecYEG channel (Tsukazaki *et al.*, 2011). This is supported by analysis of the SecDF 3D structure by electron tomography and single particle reconstruction which showed the SecDF complex adopting two different conformations (Mio *et al.*, 2014).

1.5.6 Determinants for topology of membrane proteins

The correct orientation of a protein in the membrane is reliant on the interplay of a number of topology rules and determinants including; hydrophobicity, charge difference within a helix and the positive inside rule. These are illustrated in Figure 1.9 and described below.

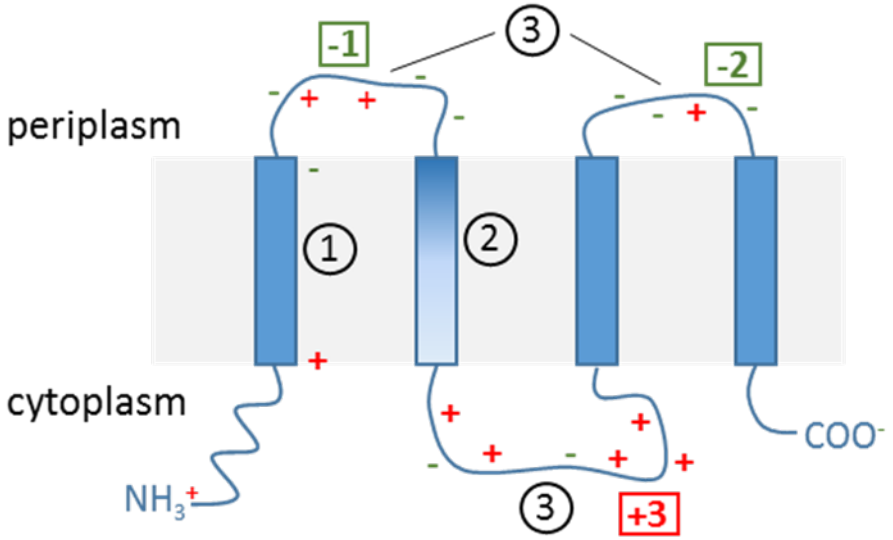


Figure 1.9 Topological determinants for the Sec machinery to correctly insert TMDs.
1) Hydrophobicity is the driving force for TMD insertion into the membrane 2) The charge difference within a TMD is based on the location of charged residues within the helix 3) The positive inside rule specifies that positive charges are more likely to be located in the cytoplasmic loop regions.

The insertion efficiency of TMDs critically depends on their hydrophobicity and length as this impacts the energetically favourable partitioning of the TMDs into the lipid bilayer (Hessa *et al.*, 2007, Hessa *et al.*, 2005). Typically this is additive, the more hydrophobic and longer the segment is then the higher the insertion propensity of the TMD (White & von Heijne, 2008). There have been a number of biological and computational hydrophobicity scales to assess the apparent free energy of insertion of a TMD into the membrane (MacCallum & Tieleman, 2011). Hessa *et al.* (2005 and 2007) investigated the insertion of engineered polypeptide segments into the endoplasmic reticulum (ER) Sec61 translocon, thereby determining the effect of residue position on TMD insertion experimentally. Together, this work has enabled the generation of a position-specific biological hydrophobicity scale for the ER membrane.

This scale shows that the energetic cost of insertion is increased if charged or polar residues are located in the central half of a TMD but this can be counterbalanced by the presence of flanking apolar residues. Also, that the interaction of the TMD with the lipid bilayer plays a role in the insertion of membrane proteins and their topology (Hessa *et al.*, 2005, Hessa *et al.*, 2007, White & von Heijne, 2008).

A similar position-specific biological hydrophobicity scale was subsequently generated for the SecYEG-mediated insertion of TMDs into the bacterial IM (Ojemalm *et al.*, 2013). The positional relevance of residues in a TMD within the bacterial IM correlates well with the values for the mammalian ER membrane (Hessa *et al.*, 2005, Hessa *et al.*, 2007). Although, there were differences such as the hydrophobicity threshold for TMD insertion is lower in the bacterial IM.

Marginally hydrophobic helices lack sufficient hydrophobicity to overcome the hydrophobicity threshold of a membrane, instead they rely on sequence-extrinsic features for their Sec-mediated recognition and insertion. Cytosolic flanking loop regions containing positive charges increase the insertion efficiency at a distance and density-specific manner (Lerch-Bader *et al.*, 2008). Other features that reduce the energetic cost of insertion include electrostatic interactions, hydrogen bond formation between neighbouring TMDs and hydrophobic packing (Zhang *et al.*, 2007, Meindl-Beinker *et al.*, 2006, Ota *et al.*, 2000).

Positive charges are four times more prevalent in the cytoplasmic regions of a membrane protein compared to the periplasmic regions, this phenomenon is known as the positive inside rule (Heijne, 1986). This is highly preserved in bacteria and consequently it is rare that many positive charges are located in the periplasmic parts of membrane proteins. In most cases it is the net positively charged domain which has an impact on the orientation of the protein (Andersson *et al.*, 1992). The explanations for the positive inside rule include that positive charges on nascent chains in the cytoplasm may interact with the negatively charged lipid head groups, therefore, remaining within the cytoplasm (Bogdanov *et al.*, 2014). Alternatively, the Sec translocon could be less accommodating

for positive charges to be translocated due to the location of oppositely charged residues compared to the positive inside rule within the translocation channel (Goder *et al.*, 2004, Monne *et al.*, 2005). The PMF may also be involved as, in *E. coli,* if the membrane potential is dissipated, negative and positive charges are rendered topologically equivalent (Andersson & von Heijne, 1994).

The positive charge bias between cytosolic and external domains correlates with the membrane topology of the protein (von Heijne, 1989, Nilsson & von Heijne, 1990). If positive charges are located in the cytosolic domains then co-translational insertion of the TMD is supported (Lerch-Bader *et al.*, 2008, Enquist *et al.*, 2009). Alternatively, charges located in non-cytosolic domains can prohibit insertion of the TMD leading to topologically frustrated helices (Gafvelin & von Heijne, 1994).

The charge difference rule gives positive and negative charges equal topological strength if they are flanking the TMDs. Therefore, the alteration of the charge balance of these residues can affect the orientation of the TMD. This has been demonstrated in yeast (Harley *et al.*, 1998) and *E. coli* (Kim et al., 1994) using single spanning hybrid chimeric constructs showing that a monotopic transmembrane protein can be engineered to adopt different orientations depending on the charge balance around the TMD.

These topological determinants are especially important in the insertion of the first TMD as this topology can affect how all the other TMDs are oriented as most proteins adopt alternating orientations relative to the plane of the membrane as a sequential mode of topogenesis (Blobel, 1980). Signal sequences are inserted into the membrane in an N_{in}-C_{out} orientation this enables cleavage by signal peptidases *trans* of the membrane. In comparison, a signal anchor (which could also be TMD1) can be inserted into the membrane in two orientations N_{out}-C_{in} and N_{in}-C_{out}. There are two models for signal anchor insertion into the ER membrane; the N-terminal inserts first and then flips (Goder & Spiess, 2003, Devaraneni *et al.*, 2011) or the insertion occurs as a hairpin-loop (Rapoport, 2007, Shaw *et al.*, 1988). The orientation of the signal anchor is dependent

on its flanking charges, hydrophobicity and length of h-domain. The main driving force for the orientation is the positive inside rule, indicating that a positive dominance at one end favours a cytosolic location (Heijne, 1986, von Heijne, 1989). The other is hydrophobicity; increased hydrophobicity slows the rate of insertion, probably due to energetically favourable interactions with the lipid bilayer through the Sec lateral gate (Higy *et al.*, 2004) and so the more hydrophobic the signal anchor the higher the probability for N_{out}-C_{in} orientation (Goder & Spiess, 2003, Wahlberg & Spiess, 1997).

Devaraneni *et al.* (2011) predicted four mechanically distinct steps for the insertion of an N_{in}-C_{out} signal anchored protein into the ER membrane. The emerging polypeptide from the ribosome is inserted in an N_{out}-C_{in} orientation, this is followed by accumulation of the nascent chain within the ribosome-Sec61 complex. A topological inversion occurs to flip the signal anchor to an N_{in}-C_{out} orientation and the C-terminal domain is then translocated into the lumen. This inversion only occurs if the positive charges at the N-terminus are sufficient whereas high hydrophobicity of the anchor prevents this inversion (Goder & Spiess, 2003).

1.6 YidC-dependent protein insertion

The YidC/Alb3/Oxa1 family of membrane proteins from the bacterial IM, the chloroplast thylakoid membrane and the mitochondrial membrane, respectively, are involved in the folding or insertion of substrates into membranes (Hennon *et al.*, 2015). *E. coli* YidC is 60 kDa with 6TMDs and a large periplasmic loop between TMD1 and TMD2 (Saaf *et al.*, 1998). YidC has 5 conserved TMD that are critical for function, however, the large periplasmic domain of *E. coli* YidC is not involved in the insertase function (Jiang *et al.*, 2003). YidC is essential in *E. coli* and a depletion of YidC results in an inhibition of membrane protein insertion (Samuelson *et al.*, 2000).

YidC has been found to crosslink to the Sec substrate (FtsQ) during its membrane insertion and is co-purified with SecYEG and SecDF-YajC (Scotti *et al.*, 2000). This indicates a role for YidC as a protein insertase cooperating with the Sec machinery. When involved in Sec-dependent protein insertion, YidC forms a supercomplex with the Sec machinery and functions as a holoenzyme (Scotti *et al.*, 2000, Schulze *et al.*, 2014). This positions YidC in close proximity to the SecY channel allowing access to proteins escaping into the lipid bilayer by the lateral gate (Sachelaru *et al.*, 2013). Therefore, YidC promotes the removal of TMD segments from the SecY channel, facilitates their integration into the lipid bilayer and acts as an assembly site for multi-spanning TMD proteins (Zhu *et al.*, 2012, Urbanus *et al.*, 2001, Beck *et al.*, 2001).

In bacteria and mitochondria, YidC and Oxa1, primarily insert and assemble protein complexes that are functionally involved in respiration (van der Laan *et al.*, 2005). In bacteria, this can occur by YidC alone or with the Sec machinery (Dalbey *et al.*, 2014). Important determinants for YidC substrates are moderately hydrophobic TMDs and charged residues in the periplasmic domains (Zhu *et al.*, 2013).

YidC has been crystallised from *Bacillus halodurans* (Kumazaki *et al.*, 2014c) and *E. coli* (Kumazaki *et al.*, 2014b). These two structures are similar and the transmembrane segments of *E. coli* YidC are shown in Figure 1.10. The conserved 5TMDs of YidC create a hydrophilic cavity which is closed from the extracellular side of the membrane but open to the cytoplasm and the lipid bilayer. At the entrance to the cavity there is a helical hairpin and within the cavity a strictly conserved positively charged residue. The C-terminal half of the helical hairpin is conserved and essential for activity in *E. coli* but the positive charge is not important in the Gram-negative YidC (Chen *et al.*, 2014). Most substrate interactions occur in the hydrophilic grove between TMD3 and TMD5 (Kumazaki *et al.*, 2014a, Yu *et al.*, 2008, Klenner & Kuhn, 2012). It is thought that the residues facing the hydrophilic cavity bind to the hydrophilic region of the substrate whereas the residues that face the membrane help to insert the hydrophobic segment into the lipid bilayer.

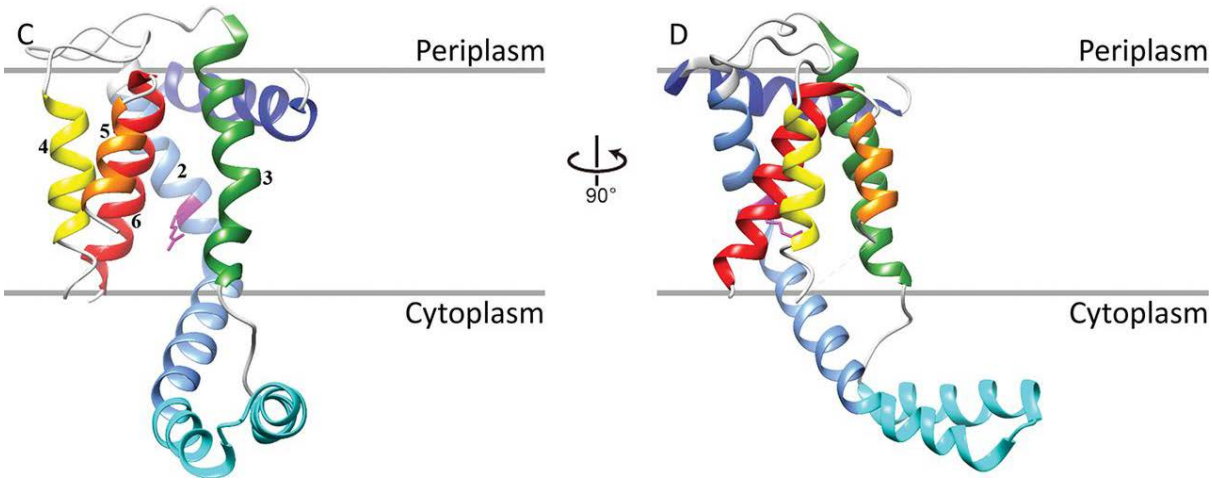


Figure 1.10 Crystal structure of the membrane portion of *E. coli* YidC.
Ribbon representations of *E. coli* YidC with the P1 domain removed and the conserved Arginine-366 shown in pink within the hydrophilic cavity. Taken from Hennon *et al.* (2015).

From the recent structures of YidC a mechanism for insertion of membrane proteins can be deduced (Kumazaki *et al.*, 2014a), the schematic is shown in Figure 1.11. Firstly the YidC substrate binds to the membrane, then the N-tail is recruited to the YidC hydrophilic groove where it potentially interacts with the conserved positive charge. The N-tail crosses the membrane which could be catalysed by the action of the PMF on the negative charges present in the translocated region and the hydrophobic interaction between the TMD and YidC. Following this, the hydrophobic region probably moves as a greasy slide between TMD3 and TMD5 to generate a TMD which is then released from YidC into the lipid bilayer.

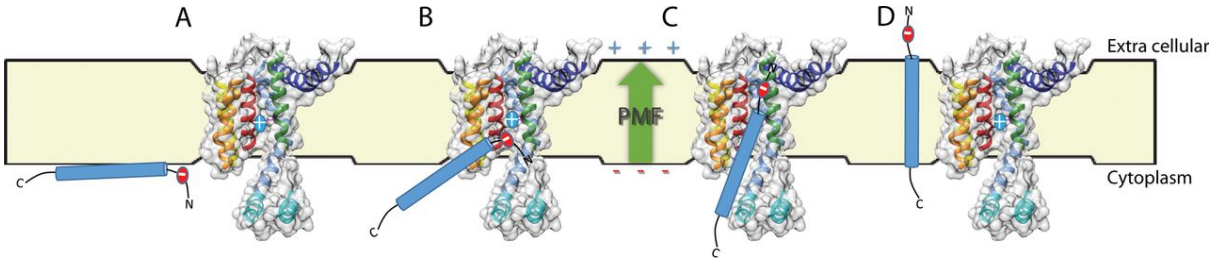


Figure 1.11 Model of membrane protein insertion by YidC.
A) The substrate is bound to the membrane through its hydrophobicity and YidC is in its resting state. B) The hydrophilic segment on the protein contacts the hydrophilic groove of YidC. C) The PMF, hydrophobic interactions between the substrate and YidC, and thinning of the membrane facilitate the translocation of the substrate. D) The substrate is released into the lipid bilayer and YidC returns to resting state. Taken from Hennon *et al.* (2015).

1.7 The twin-arginine (Tat) pathway

The first insight into a Sec-independent translocation pathway which transports folded proteins was from study of plant thylakoid membranes. It was found that a proton gradient (ΔpH) was required for translocation of two lumenal proteins across the thylakoid membrane (Mould & Robinson, 1991). This ΔpH-dependent pathway was recognised to have the ability to translocate fully-folded globular proteins, this is in contrast to the Sec pathway which translocates unfolded proteins (Clark & Theg, 1997, Hynds *et al.*, 1998). The targeting of these folded proteins is dependent on a signal peptide that contains a targeting motif which comprises twin-arginine residues that are essential for translocation and when mutated the transport of known substrates is prevented (Chaddock *et al.*, 1995). A mutation, *hcf106*, disrupting the translocation of proteins through the ΔpH-dependent pathway was identified in maize (Settles *et al.*, 1997). This *hcf106* gene shows homology to bacterial genes indicating a conservation of this pathway between bacteria and plants.

Similar twin-arginine motifs were located in the signal sequences of a number of bacterial proteins which contain oxygen sensitive complex cofactors (Berks, 1996). It was hypothesised that these proteins would fold in the cytoplasm as this was the compartment in which the cofactors would interact with their proteins. Therefore, a different translocating pathway was required as the SecYEG translocon would be unable to transport these folded proteins across the IM. An operon *mttABC* was identified, which was required for Sec-independent translocation of proteins presenting the twin-arginine motif in their signal peptide and harbouring a cofactor-containing globular domain (Weiner *et al.*, 1998). The mutation *mttA* had no effect on the translocation of Sec substrates, however, it prevented the periplasmic location of NapA (nitrate reductase), TorA (trimethylamine N-oxide reductase) and DmsA (subunit of the dimethylsulfoxide reductase). Subsequently, two *E. coli* genes were identified which were homologous to the maize *hcf106* genes and were required for the export of cofactor-containing proteins with a twin-arginine motif in their signal peptide (Sargent *et al.*, 1998). This novel pathway

which exports folded proteins was named after the targeting motif; the twin-arginine translocase (Tat) pathway.

1.7.1 The Tat signal peptide

Similar to Sec-dependent proteins, substrates of the Tat pathway contain an N-terminal cleavable signal peptide used for targeting to the translocation machinery (Berks, 1996). They have a related tripartite structure consisting of a polar n-region, a hydrophobic h-region and a basic c-region (Figure 1.12), but there are some critical differences compared to the Sec signal peptides. The most distinguishable is the Tat motif which is essential for targeting and it is located in the junction of the n- and h-regions of the signal peptide.

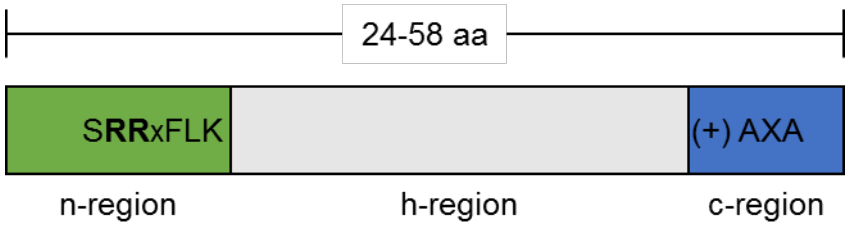


Figure 1.12 The tripartite structure of the Tat signal peptide.
The N-terminal Tat signal peptide has a tripartite structure; a basic n-region (green), a hydrophobic h-region (grey) and a polar c-region (blue). The essential Tat motif is found within the n-region, whereas the signal peptidase cleavage site is located within the c-region.

This motif, in bacteria, is well conserved with a consensus sequence; S**RR**xLFK, where the twin-arginines are usually invariant (Berks, 1996). There have been many studies investigating the requirement of these residues for transport. Site-directed mutagenesis of the signal peptide of the *E. coli* Tat substrates SufI and YacK/CueO found that substitutions of one arginine to a positively charged lysine maintained slow transport through Tat but substitutions to twin-lysine inhibited Tat translocation (Stanley *et al.*, 2000). The second arginine is predicted to be more critical for Tat transport (Buchanan *et al.*, 2001). Indeed, there are several natural Tat substrates that have variant signal peptides where the first arginine is replaced (Hinsley *et al.*, 2001, Ignatova *et al.*, 2002). Although the twin-arginine motif is the hallmark for targeting to the Tat machinery alone,

they do not prevent translocation by the Sec machinery as there are several Sec targeting substrates that also contain twin-arginines in their signal sequences.

The residue preceding the twin-arginine motif, located position -1 relative to the conserved arginines, is frequently a serine and this residue is predicted to act as a helix capping residue and stabilises a potential α-helix that is formed by the signal peptide. However, when this serine residue in the signal peptide of SufI was substituted to an alanine no significant effect was found (Stanley *et al.*, 2000), this is in contrast to similar substitutions in the signal peptides of TorA and DmsD which showed either a severe slowing or inhibition of export (Mendel *et al.*, 2008).

The residue at position +3, directly following the twin-arginines, is not conserved, therefore, is not predicted to have a role in Tat transport. However, this residue is predicted to be required as a spacer between the twin-arginine motif and the following hydrophobic residues (Buchanan *et al.*, 2001, Stanley *et al.*, 2000).

The highly conserved phenylalanine, located at position +4, was found to be important for transport. Substitution of this residue for a less hydrophobic residue (tyrosine) or charged residues (arginine or aspartate) severely decreased or entirely abolished Tat translocation (Stanley *et al.*, 2000, Mendel *et al.*, 2008). Hydrophobicity also seems important for the consensus leucine (position +5). In contrast, the positively charged lysine residue in position +6 that separates the hydrophobic residues from the h-region is thought to slow down the translocation of substrates, potentially to allow time for cofactor insertion (Stanley *et al.*, 2000).

The h-regions of Tat signal peptides (Figure 1.12) are usually less hydrophobic than Sec signal peptides. It has been found that increasing the hydrophobicity of the h-region of the signal peptide of TorA which was fused to the P2 loop of the Sec-dependent IM protein Leader peptidase (Lep) rerouted translocation to the Sec machinery (Cristobal *et al.*, 1999). This indicates that the hydrophobicity of the h-region plays a role in Sec avoidance. The positive charges located in the c-region have also been found to be a Sec avoidance motif critical in preventing targeting to the Sec machinery (Cristobal *et*

al., 1999, Bogsch *et al.*, 1997, Blaudeck *et al.*, 2003). The c-region also contains the signal peptidase recognition site, this is usually an AxA motif which allows cleavage of the signal peptide by LepB (Luke *et al.*, 2009). This presumably occurs in the later stages of translocation when the c-region of the signal peptide reaches the periplasm.

1.7.2 Tat substrates

The number of Tat substrates which have been predicted and experimentally confirmed varies between organisms. *E. coli* currently has 27 experimentally confirmed substrates which are exported into the periplasm using a twin-arginine signal peptide (Tullman-Ercek *et al.*, 2007, Palmer & Berks, 2012, Berks *et al.*, 2005). A list of current substrates is shown in Table 1.1, this reveals that the substrates which utilise the Tat pathway usually contain a number of different metal cofactors. Although the Tat pathway is not essential in *E. coli* many of these substrates are involved in important bacterial processes which are affected in *tat* mutant strains; for example, cell growth, motility and iron acquisition.

Table 1.1 List of known and predicted Tat substrates of *E. coli* K-12.
In some cases a Tat substrate lacks its own signal peptide and so is exported in a complex with a signal peptide-binding partner. These substrates listed have the name of the partner and the sequence of the signal peptide used for export. Assembly chaperones, which are often involved in the insertion of bound cofactors, are also indicated. ? indicates unknown cofactors (Palmer *et al.*, 2010, Berks *et al.*, 2005).

Protein	Physiological role	Cofactors	Potential Tat chaperone(s)	Signal peptide	Sequence of signal peptide
Proteins with complex multiatom cofactors					
DmsA	DMSO reductase α-subunit	MGD, [4Fe-4S]	DmsD	DmsA	MKTKIPDAVLAAEVSRRGLVKTTAIGGLAMASSALTLPFSRIAHA
DmsB	DMSO reductase β-subunit	4 x [4Fe-4S]	DmsD	DmsA	MKTKIPDAVLAAEVSRRGLVKTTAIGGLAMASSALTLPFSRIAHA
FdnG	Nitrate-inducible formate dehydrogenase α-subunit	MGD, [4Fe-4S]	FdhE	FdnG	MDVSRRQFFKICAGGMAGTTVAALGFAPKQALAQA
FdnH	Nitrate-inducible formate dehydrogenase β-subunit	4 x [4Fe-4S]	FdhE	FdnG	MDVSRRQFFKICAGGMAGTTVAALGFAPKQALAQA
FdoG	Formate dehydrogenase-O α-subunit	MGD, [4Fe-4S]	FdhE	FdoG	MQVSRRQFFKICAGGMAGTTAAALGFAPSVALA
FdoH	Formate dehydrogenase-O β-subunit	4 x [4Fe-4S]	FdhE	FdoG	MQVSRRQFFKICAGGMAGTTAAALGFAPSVALA
HyaA	Hydrogenase-1 small subunit	2 x 4Fe-4S]	HyaE, HyaF	HyaA	MNNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMAPKIAWA[
HyaB	Hydrogenase-1 large subunit	[Ni-Fe$(CN)_2$CO]	HyaE, HyaF	HyaA	MNNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMAPKIAWA[
HybA	Electron transfer to hydrogenase 2	4 x [4Fe-4S]	-	HybA	MNRRNFIKAASCGALLTGALPSVSHA
HybC	Hydrogenase-2 large subunit	[Ni-Fe$(CN)_2$CO]	HybE	HybO	MTGDNTLIHSHGINRRDFMKLCAALAATMGLSSKAAA
HybO	Hydrogenase-2 small subunit	2 x [4Fe-4S]	HybE	HybO	MTGDNTLIHSHGINRRDFMKLCAALAATMGLSSKAAA
NapA	Periplasmic nitrate reductase	MGD, [4Fe-4S]	NapD	NapA	MKLSRRSFMKANAVAAAAAAAGLSVPGVA
NapG	Electron transfer to periplasmic nitrate reductase	4 x [4Fe-4S]	-	NapG	MSRSAKPQNGRRRFLRDVVRTAGGLAAVGVALGLQQQTARA
NrfC	Electron transfer to periplasmic nitrate reductase	4 x [4Fe-4S]	-	NrfC	MTWSRRQFLTGVGVLAAVSGTAGRVVA
TorA	TMAO reductase	MGD	TorD	TorA	MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAAQA
TorZ	TorA homologue	MGD	?	TorZ	MTLTRREFIKHSGIAAGALVVTSAAPLPAWA

YagR	Xanthine dehydrogenase family domain homologue	MCD?	YagQ	YagT	MSNQGEYPEDNRVGKHEPHDLSLTRRDLIKVSAATAVVYPH STLAASVPA
YagS	Xanthine dehydrogenase family domain homologue	FAD	YagQ	YagT	MSNQGEYPEDNRVGKHEPHDLSLTRRDLIKVSAATAVVYPH STLAASVPA
YagT	Xanthine dehydrogenase family domain homologue	2 x [4Fe-4S]	YagQ	YagT	MSNQGEYPEDNRVGKHEPHDLSLTRRDLIKVSAATAVVYPH STLAASVPA
YdhX	Iron-sulfur protein	4 x [4Fe-4S]	-	YdhX	MSWIGWTVAATALGDNQMSFTRRKFVLGMGTVIFFTGSASS LLA
YedY	Sulfite oxidase superfamily homologue	Molybdopterin	-	YedY	MKRRQVLKALGISATALSLPHAAHA
YnfE	DmsA homologue	MGD, [4Fe-4S]	DmsD	YnfE	MSKNERMVGISRRTLVKSTAIGSLALAAGGFSLPFTLRNAAA
YnfF	DmsA homologue	MGD, [4Fe-4S]	DmsD	YnfF	MMKIHTTEALMKAEISRRSLMKTSALGSLALASSAFTLPFSQ MVRA
YnfG	DmsB homologue	4 x [4Fe-4S]	DmsD	YnfE/YnfF	MSKNERMVGISRRTLVKSTAIGSLALAAGGFSLPFTLRNAAA
Proteins that lack complex multiatom cofactors					
AmiA	N-acetylmuramoyl-$_L$-alanine amidase	Metal ions?	-	AmiA	MSTFKPLKTLTSRRQVLKAGLAALTLSGMSQAIA
AmiC	N-acetylmuramoyl-$_L$-alanine amidase	Metal ions?	-	AmiC	MSGSNTAISRRRLLQGAGAMWLLSVSQVSLA
CueO	Multicopper oxidase linked to copper resistance	Cu ions	-	CueO	MQRRDFLKYSVALGVASALPLWSRAVFA
FhuD	Ferrichrome-binding protein	Fe(III)	-	FhuD	MSGLPLISRRRLLTAMALSPLLWQMNTAHA
MdoD	Involved in osmoregulated periplasmic glucan biosynthesis	?	-	MdoO	MDRRRFIKGSMAMAAVCGTSGIASLFSQAAFA
SufI	Multicopy suppressor of an *ftsI* mutation	None	-	SufI	MSLSRRQFIQASGIALCAGAVPLKASA
YaeI	Probable metallo-phosphoesterase	Metal ions?	-	YaeI	MISRRRFLQATAATIATSSGFGYMHYC
YahJ	Probable metallo-hydrolase	Metal ions?	-	YahJ	MKESNSRREFLSQSGKMVTAAALFGTSVPLAHA
YcbK	Unknown	?	-	YcbK	MDKFDANRRKLLALGGVALGAAILPTPAFA
YcdB	Unknown; structural gene overlaps that of YcdO	?	-	YcdB	MQYKDENGVNEPSRRRLLKVIGALALAGSCPVAHA
YcdO	Unknown; structural gene overlaps that of YcdB	?	-	YcdO	MTINFRRNALQLSVAALFSSAFMANA
WcaM	Coded in colonic acid biosynthetic gene cluster; possible lyase	?	-	WcaM	MPFKKLSRRTFLTASSALAFLHTPFARA

Streptomyces species are predicted to have a lot more Tat substrates than *E. coli*, indicating that this translocation pathway is more important in these bacteria. For example, bioinformatic analysis of the genome sequence of *S. coelicolor* suggested that 189 proteins may be substrates of the Tat pathway (Dilks *et al.*, 2003, Bendtsen *et al.*, 2005). Proteomic studies of *S. coelicolor* strains with and without the Tat system along with testing of candidate Tat signal peptides in Tat-dependent reporter assays has confirmed 33 Tat substrates (Widdick *et al.*, 2006, Li *et al.*, 2005). More than 100 substrates of *S. scabies* have also been predicted to be exported using the Tat pathway and 47 substrates have been verified using an agarose reporter assay with candidate Tat signal peptides (Joshi *et al.*, 2010).

In comparison to most bacteria, *M. tuberculosis* has an essential functional Tat system (Palmer & Berks, 2012, Saint-Joanis *et al.*, 2006). 18 Tat substrates have been confirmed using a *Bla* reporter fused to the candidate Tat signal sequence (Ligon *et al.*, 2012, McDonough *et al.*, 2005). A number of these substrates are known to contribute directly to virulence and drug resistance, for example, the β-lactamase BlaC and phospholipase C enzymes (McDonough *et al.*, 2005, Raynaud *et al.*, 2002). As the Tat system is absent in humans and essential in *M. tuberculosis*, it has been investigated as a potential drug target for future treatment of TB (Vasil *et al.*, 2012).

Tat substrates vary in folded size from 20 to 70 Å in diameter, therefore, the Tat translocase machinery must be able to accommodate this range of substrates while maintaining an ionic seal (Berks *et al.*, 2000). This begs the question as to the need for the existence of the Tat pathway and there are several reasons why Tat substrates are required to be folded in the cytoplasm before export.

Primarily, insertion of non-covalently bound complex metal cofactors must take place in the cytoplasm. Some of these cofactors are highly oxygen sensitive, therefore, they require insertion into the protein in a reducing environment like the cytoplasm (Berks, 1996). Insertion of other cofactors (for example, nickel into hydrogenase) requires the

hydrolysis of nucleotide triphosphates, like GTP, which are only available in the cytoplasm (Maier *et al.*, 1995).

Folding in the cytoplasm could avoid competition between different metal ions for the active site of the proteins which could result in the incorporation of an inappropriate metal ion. The cell controls this by compartmentalisation, for example, in the cyanobacterium *Synechocystis* PCC 6803 the Mn^{2+}-binding protein, MncA, will only bind the appropriate metal ion if it is in molar excess over other ions, as Mn^{2+} is a weak-binding metal (Tottey *et al.*, 2008). Therefore, this ligand binding occurs in the cytoplasm, where other competing metal ions, like Cu^{2+} and Zn^{2+} are already tightly bound to other proteins, before export into the periplasm by the Tat pathway. In comparison, CucA, which binds Cu^{2+}, is exported by the Sec pathway and folds in the periplasm where there is an abundance of unbound Cu^{2+}. This highlights the importance of the different targeting pathways to ensure the appropriate metal is bound to the protein.

There are also a number of Tat substrates that do not contain cofactors, such as, the cell wall amidases AmiA and AmiC (Ize *et al.*, 2003). Other Tat substrates include SoxY and SoxZ neither of which contain cofactors, however, they do form a heterodimer, SoxYZ, which is involved in thiosulfate oxidation in *Paracoccus pantrophus*. Only SoxY maintains a Tat signal peptide, consequently, these two proteins fold and form a complex before co-translocation through the Tat machinery with SoxY carrying SoxZ in a 'piggy-back' manner (Sauve *et al.*, 2007). Another example of this hitchhiking is *E. coli* hydrogenase 2. The small subunit (HybO) contains a Tat signal peptide whereas the large subunit (HybC) does not. However, when either of these proteins are expressed without the other they accumulate in the cytoplasm, indicating that the signal sequence on HybO is not sufficient to translocate this subunit without the folding and dimerization of HybC (Rodrigue *et al.*, 1999). This leads to an alternative justification of the presence of the Tat pathway, for the co-translocation of heterodimeric complexes.

The Tat pathway is versatile and has also been found to translocate lipoproteins and some membrane-bound substrates. Several small subunits of bacterial respiratory

formate dehydrogenases and uptake hydrogenases are membrane anchored by a single carboxy-terminal helix (C-tail). These include; hydrogenase-2 subunits HybO and HybA, hydrogenase-1 subunit HyaA, nitrate inducible formate dehydrogenase subunit FdnH and formate dehydrogenase-O β-subunit FdoH (Hatzixanthis *et al.*, 2003). The *E. coli* FdnH, a Tat substrate and a major component of nitrate respiration, has been shown to contain a TMD by crystallography (Jormakka *et al.*, 2002). This C-tail is potentially a 'stop-transfer' sequence although it could also be laterally translocated into the membrane. The use of a YidC depletion strain showed that YidC is not involved with the translocation or insertion of these membrane bound Tat substrates (Hatzixanthis *et al.*, 2003). Another known membrane-bound substrate is the Rieske Iron-sulfur (FeS) protein. In contrast, this Tat substrate is anchored to the membrane *via* an N-terminal helix composed of its Tat signal peptide (Molik *et al.*, 2001, Aldridge *et al.*, 2008, Bachmann *et al.*, 2006, De Buck *et al.*, 2007).

Although non-essential in most organisms, the Tat pathway is present in many bacterial pathogens, for example, the opportunistic pathogen *Pseudomonas aeruginosa* (Voulhoux et al., 2001), the enterohaemorrhagic *E. coli* serotype O157:H7 (Pradel et al., 2003) and *Legionella pneumophila* (Rossier & Cianciotto, 2005). Conversely, the Tat machinery is essential for viability in several organisms; *M. tuberculosis* (Saint-Joanis et al., 2006), *haloarchaea* (Dilks et al., 2005), *Sinorhizobium melitoti* (Pickering & Oresnik, 2010) and *Bdellovibrio bacteiovorus* (Chang et al., 2011). This indicates important roles for Tat substrates and the Tat pathway in virulence.

1.7.3 Substrate proofreading and Quality control

The Tat system has been found to some extent to distinguish between folded and unfolded substrates. A fusion of a Tat signal peptide to the Sec-dependent PhoA showed no transport into the periplasm as PhoA remains unfolded in the cytoplasm due to the reducing environment which prevents the formation of two intra-molecular disulphide bonds that are essential for activity and stability (Stanley *et al.*, 2002). However, the same fusion expressed in cells engineered to have an oxidising cytoplasm allowed the folding of the protein and therefore subsequent translocation through the Tat pathway (DeLisa *et al.*, 2003). This led to the hypothesis of a folding quality control mechanism.

Nevertheless, the *E. coli* Tat machinery has also been found to translocate small unstructured hydrophilic proteins into the periplasm. Unfolded high potential iron-sulfur protein (HiPIP) was found to be compatible with translocation through the Tat system as it is sufficiently hydrophilic in an unfolded state, however, the addition of a stretch of hydrophobic residues abolished translocation (Richter *et al.*, 2007). This is likely due to the interactions of the exposed hydrophobic residues with the membrane at a late stage in translocation, whereas, in folded proteins these residues are usually hidden. Therefore, it is unlikely that the Tat machinery has a specific quality control mechanism, but exposed hydrophobic segments in unfolded proteins are prevented from translocating by interacting with the lipids in the membrane.

Proof reading chaperones are not essential for all Tat substrates, as an *in vitro* assay found that SufI was able to be translocated in the absence of any cytosolic chaperones (Holzapfel *et al.*, 2009). Nevertheless, chaperone proteins have been found to interact with certain Tat substrates before translocation, as depicted in Figure 1.13.

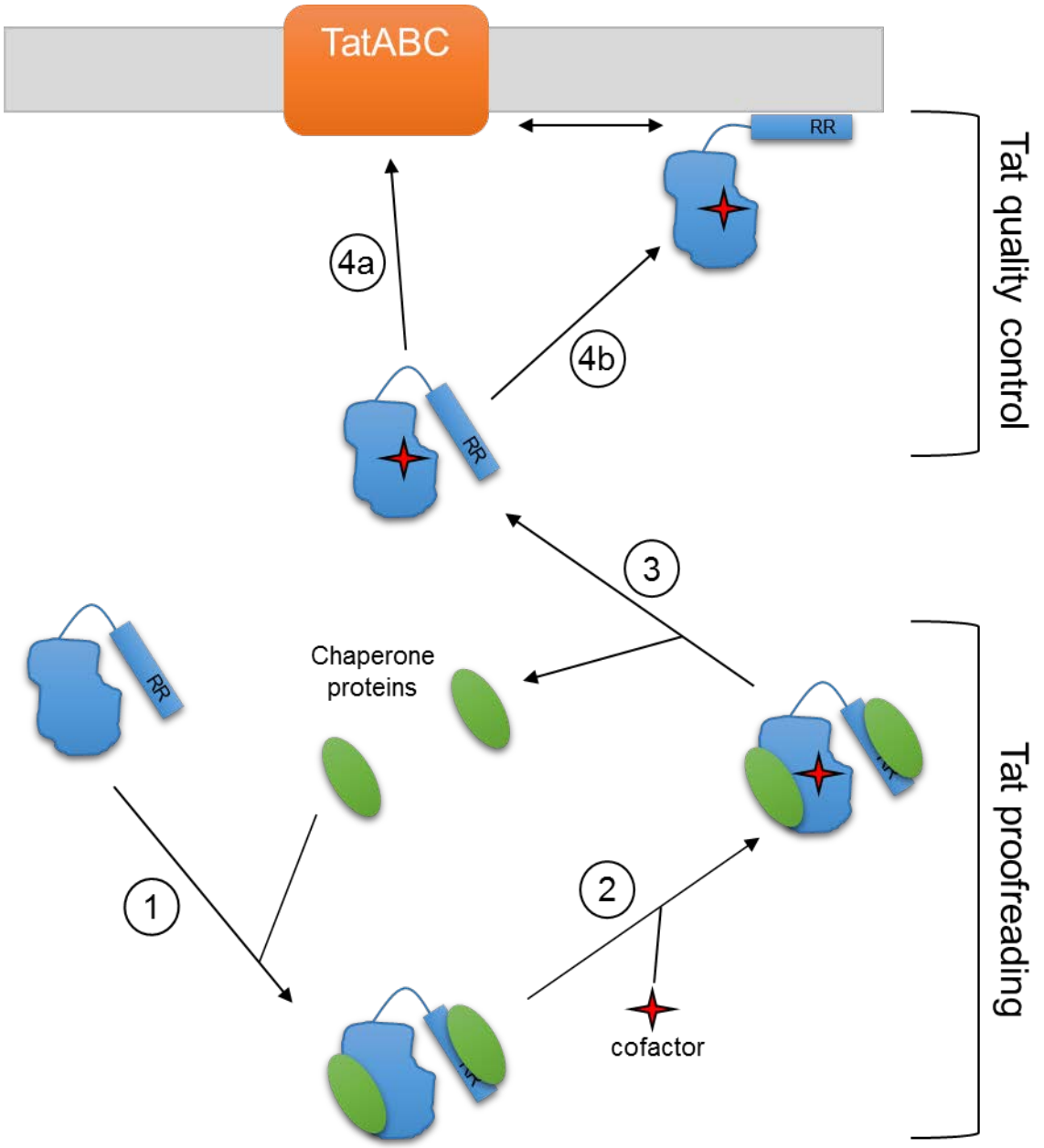


Figure 1.13 Quality control and proofreading in the Tat system.
The correct insertion of cofactors and folding of the Tat substrate protein is ensured by Tat proofreading and quality control.1) Chaperones (green) are able to tightly bind to the signal peptide (RR) or domains within the apoprotein (blue). 2) Cofactor (red) insertion into the substrate and binding of any partner subunits to form a complex (not shown). 3) Chaperones are released and Tat substrates either 4a) interact directly with the Tat system or 4b) bind to the lipid bilayer before lateral diffusion toward the Tat system (Bageshwar *et al.*, 2009, Shanmugham *et al.*, 2006). The Tat system has been proposed to reject unfolded Tat substrates in a process called quality control (DeLisa *et al.*, 2003). Figure adapted from Palmer *et al.* (2010).

DmsD is essential for the biogenesis of the molybdopterin guanine dinucleotide (MGD)-containing *E. coli* enzyme dimethyl sulfoxide (DMSO) reductase (DmsA) and in *dmsD* mutant strains there is no DMSO reductase activity. DmsA binds MGD as a catalytic cofactor and DmsD interacts with this preprotein through its signal peptide (Oresnik *et al.*, 2001).

Another chaperone, TorD, which is closely related to DmsD (Sargent *et al.*, 2002), is involved in the biogenesis of trimethylamine N-oxide (TMAO) reductase (TorA). TorD binds to both the signal peptide and the unfolded mature domain of TorA before the molybdenum cofactor has been inserted (Jack *et al.*, 2004, Hatzixanthis *et al.*, 2005, Pommier *et al.*, 1998). Recently, a pre-export cofactor-free TorAD complex was characterised using small-angle X-ray scattering (SAXS) finding a 1:1 stoichiometry which identified two high-affinity binding sites for TorD. When the Tat signal sequence of TorA was removed, a stable TorAD complex was still able to be isolated suggesting that the correct folding of TorA when the cofactor is inserted, triggers its release from TorD prior to translocation through the Tat machinery (Dow *et al.*, 2013).

NapD, is another Tat chaperone unrelated to TorD and DmsD, and is a small monomeric cytoplasmic protein which is essential for the biogenesis of the periplasmic nitrate reductase (NapA). NapD interacts with primarily with the NapA signal peptide (Maillard *et al.*, 2007). The signal peptide of NapA has been found to undergo conformational change upon binding NapD. Analytical centrifugation and SAXS have found the NapAD complex in a 1:1 stoichiometry, however, unlike TorAD, this complex cannot form when the Tat signal peptide is removed indicating that the NapD chaperone binds solely to the NapA signal peptide (Dow *et al.*, 2014). These chaperone-substrate contacts have been proposed to delay the interaction of the substrate with the Tat machinery to allow time for the passenger domain to fold around the inserted cofactor.

1.7.4 Components of the Tat translocon

The *E. coli* Tat machinery is a 600 kDa complex consisting of four related membrane proteins; TatA, TatB, TatC and TatE (Bogsch *et al.*, 1998, Sargent *et al.*, 1998, Weiner *et al.*, 1998). Three of these are encoded on one operon (Figure 1.14A), *tatA, tatB* and *tatC,* along with a fourth gene, *tatD.* TatD is a cytoplasmic protein with DNase activity but has no function in protein export by the Tat system (Wexler *et al.*, 2000, Sargent *et al.*, 1998). The *tatE* gene is monocistronic (Figure 1.14B), thought to be a cryptic gene duplication encoding a functional homologue of TatA (Sargent *et al.*, 1998, Bolhuis *et al.*, 2001).

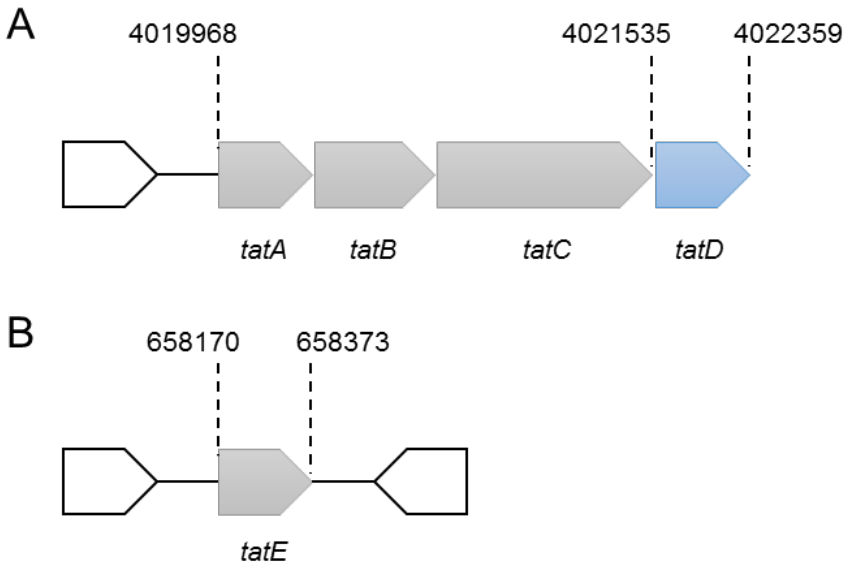


Figure 1.14 Organisation of *E. coli tat* genes.
Genes encoding components of the Tat system are specified in grey. Nearby genes are shown in white. Above the genes the base numbers on the *E. coli* MG1655 genome are indicated. A) The essential genes of the Tat system are located in a single operon alongside *tatD*, a gene that is co-transcribed but not involved with protein translocation (Wexler *et al.*, 2000). B) *tatE* gene, located as a monocistronic operon found in a different region of the genome (Sargent *et al.*, 1998).

Expression studies have shown that *tatABC* and *tatE* are constitutively expressed, indicating the importance of the translocase system for both aerobic and anaerobic growth conditions (Jack *et al.*, 2001). Nevertheless, it has been found that *tatE* is expressed at significantly lower levels than *tatA*. TatE has over 50% identity with TatA and they have overlapping functions so that one protein can partially substitute

functionally for the other, although it has been found that TatA is more important for Tat translocation (Sargent *et al.*, 1998). A cellular molar ratio has been determined for the main Tat components of TatA:B:C as 40:2:1 (Jack *et al.*, 2001, Sargent *et al.*, 1998). However, a recent *in vivo* study has found that TatE potentially interacts with active Tat translocases and thus, might be a constituent component of TatABC-based translocases under certain conditions (Eimer *et al.*, 2015).

The TatA family encompasses proteins TatA, TatB and TatE. All of these proteins are membrane spanning with one TMD at the N-terminus followed by a short hinge region and an amphipathic helix at the C-terminal lying at the membrane-water interface (Hu *et al.*, 2010). In *E. coli,* both TatA and TatB are essential for Tat transport indicating that they have distinct roles, conversely, some Gram-positive bacteria have minimal Tat systems comprising of just TatA and TatC (Freudl, 2013, Robinson *et al.*, 2011). These TatA proteins appear to be bifunctional as the *Bacillus subtilis* TatA can complement both *tatA* and *tatB* deletions in *E. coli* (Barnett et al., 2008). It seems that early gene duplication has led to TatB diverging in sequence and function from TatA giving them distinct roles in Tat translocation (Yen et al., 2002).

Solution NMR structures have been obtained for *E. coli* TatA and TatB proteins, seen in Figure 1.15 (Rodriguez *et al.*, 2013, Zhang *et al.*, 2014), and *B. subtilis* TatAd which is representative of the TatA from a minimal Tat system (Hu *et al.*, 2010). All these proteins have a similar core structure with a short hydrophobic N-terminal helix suggested to span the membrane, positioned at a right angle to an amphipathic helix (APH) which lies approximately along the membrane surface. Protease sensitivity experiments and cysteine accessibility have found that the C-terminus of these proteins is located in the cytoplasm (Porcelli *et al.*, 2002, Koch *et al.*, 2012).

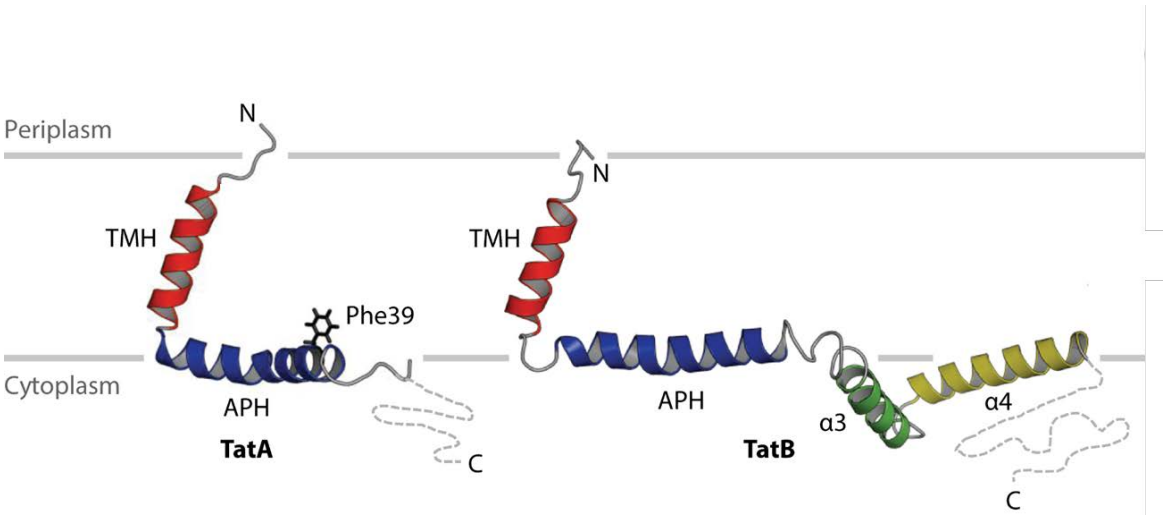


Figure 1.15 Solution NMR structures of *E. coli* TatA and TatB.
The transmembrane helices (TMHs) and amphipathic helices (APHs) are indicated and the natively unstructured C-terminal tails are represented by dashed lines. The essential residue F39 is depicted for TatA. Structures were obtained by Zhang *et al.* (2014) and Rodriguez *et al.* (2013). Adapted from Berks (2015).

TatA proteins oligomerise and increase their association with TatC in response to substrate binding and are believed to make up the translocation channel. The TMD is essential for oligomerisation (De Leeuw *et al.*, 2001, Porcelli *et al.*, 2002) and the APH is vital for function indicated by its intolerance to mutations (Hicks *et al.*, 2005, Barrett *et al.*, 2003, Barrett & Robinson, 2005). Crosslinking and single molecule florescence discovered that TatA exists in an unstimulated membrane as a homotetramer (Dabney-Smith & Cline, 2009, Leake *et al.*, 2008), whereas the formation of larger oligomers generates a ring structure which can be seen with single particle electron microscopy and is thought to be the translocation channel (Gohlke *et al.*, 2005).

TatB proteins exist in an equimolar complex with TatC and are part of the substrate recognition complex. They have a similar structure to TatA but with an additional two APH before the disordered C-terminal tail. This longer C-tail can be deleted without eliminating TatB function (Lee *et al.*, 2002). Many studies have found that the TMD of TatB associates with TatC TMD5, this is known through suppressor analysis (Kneuper *et al.*, 2012), site-directed photocrosslinking (Blummel *et al.*, 2015) and disulphide scanning crosslinking (Kneuper *et al.*, 2012, Rollauer *et al.*, 2012).

Recently structures of the TatC protein of the thermophilic bacterium *Aquifex aeolicus* have been determined by X-ray crystallography in three distinct crystal environments (Ramasamy *et al.*, 2013, Rollauer *et al.*, 2012). The similarity between these structures indicates that TatC has limited conformational flexibility. The *A. aeolicus* TatC has 40% sequence identity to the *E. coli* TatC which has allowed functional data to be mapped onto the structure giving a greater understanding of the interactions of the TatC protein. The structure, shown in Figure 1.16, indicates that the TatC TMDs create a folded shape of a cupped hand where the 'palm' is a strikingly broad concave cavity and the loops 1 and 2 act to stabilise the structure with an overhanging cap. TMD5 and TMD6 are too short to span the membrane generating a gap between the helix ends and the capping structure, thought to be a route between the cavity and the periplasm. The membrane is distorted around these two helices which thins the bilayer, potentially as part of the translocation process.

Mutagenesis studies have discovered several key residues in TatC involved in Tat translocation. A phenylalanine (F94 in *E. coli*) residue located in the first cytoplasmic loop has been found to inactivate Tat and prevent substrate binding, possibly due to the hindrance of substrate recognition (Buchanan *et al.*, 2002, Strauch & Georgiou, 2007, Holzapfel *et al.*, 2007). TatC also contains a strictly conserved glutamate/glutamine which is located in the centre of the TatC cavity thought to generate a polar environment and replacement of this residue compromises *E. coli* Tat translocation (Holzapfel *et al.*, 2007, Buchanan *et al.*, 2002).

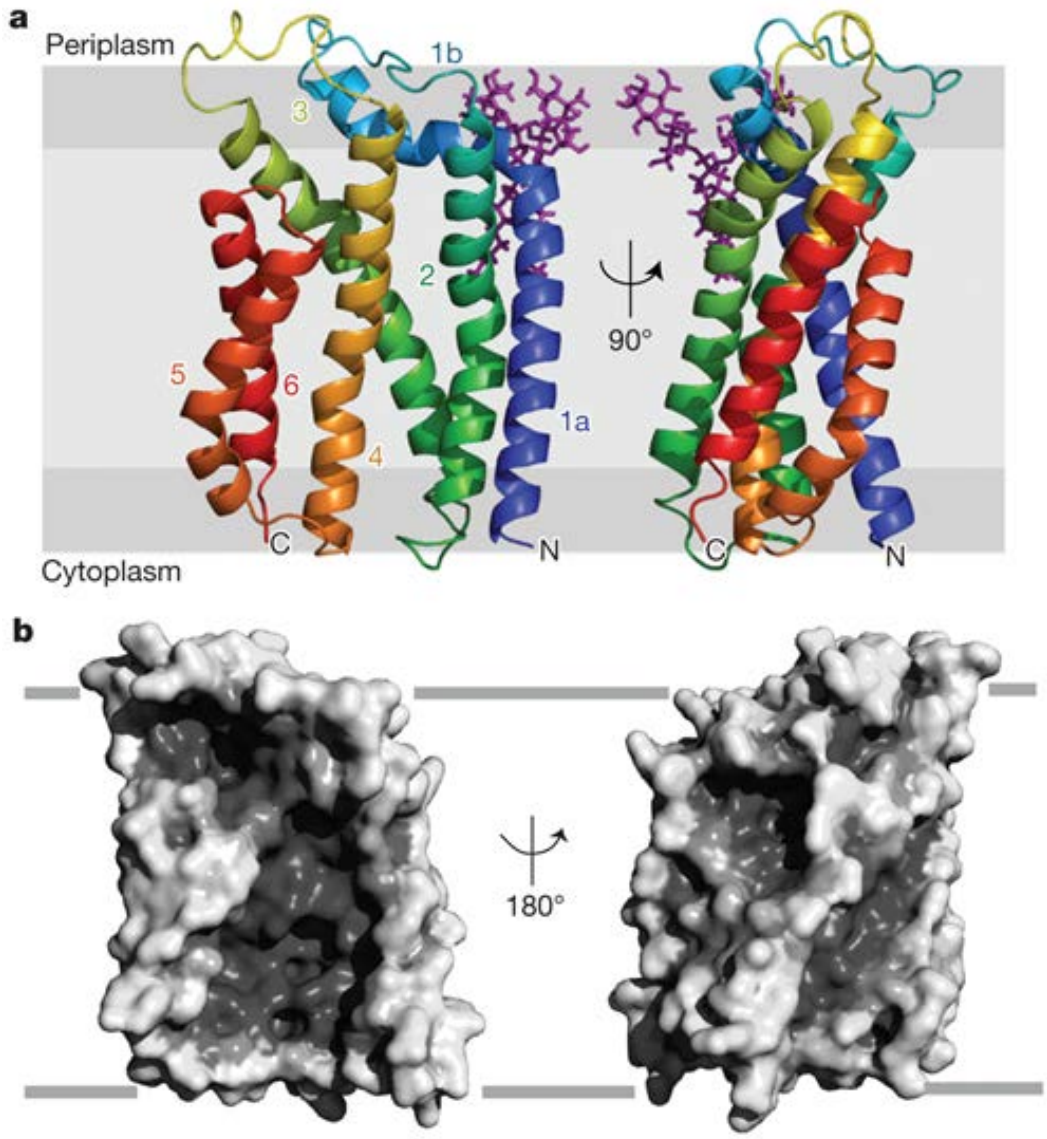


Figure 1.16 X-ray crystal structure of *A. aeolicus* TatC.
A) Cartoon representation of the TatC structure coloured from blue (N-terminus) to red (C-terminus) with the helix number in the corresponding colour. The position of the membrane bilayer is predicted by molecular dynamic simulations. A semi-ordered lauryl maltose neopentyl glycol (LMNG) detergent molecule is shown in purple. B) Surface representation of the TatC model. Left and right hand views in A and B are the same orientation. Taken from Rollauer *et al.* (2012).

The TatBC complex is involved in the recognition of the substrate which is the initiating step for Tat translocation. There are 6-8 copies of TatB and TatC proteins in this complex with equal stoichiometry of each protein (Bolhuis *et al.*, 2001, Tarry *et al.*, 2009). Crosslinking experiments have shown that the signal peptide of substrates interacts with TatC whereas the passenger domain and the h-region of the signal peptide interacts with TatB (Alami *et al.*, 2003, Gerard & Cline, 2006, Maurer *et al.*, 2010). The binding of the

twin-arginine motif to TatC mediates specificity of substrate recognition (Rollauer *et al.*, 2012, Zoufaly *et al.*, 2012, Ma & Cline, 2010, Kreutzenbeck *et al.*, 2007, Strauch & Georgiou, 2007). It has been found that the n-region of the signal peptide interacts with a surface patch on the cytoplasmic face of TatC, comprising the N-terminal tail and the loop region between TMD2 and TMD3. The twin-arginines interact with two conserved glutamic acid residues, whereas the consensus phenylalanine in the signal peptide is proposed to stack with a conserved phenylalanine (F94 in *E. coli*) on TatC (Berks *et al.*, 2014).

Gerard *et al.* (2007) studied Thylakoid Tat translocation and discovered that the signal peptide can interact with the TatBC complex in two different modes; peripheral-binding and deep-interaction. The peripheral-binding mode permitted weak electrostatic interactions between the signal peptide and the TatBC complex allowing proteolysis of the signal peptide to occur. The deep-interaction mode enabled the signal peptide to be buried within the TatBC complex, therefore, inaccessible for proteolysis. Substrates could convert to deep-interaction mode when the PMF was present (Gerard & Cline, 2007). Similarly, two binding modes have been recognised for an *E. coli* Tat substrate, weak-binding to the TatBC complex where the substrate can be exchanged and strong-binding where the substrate and TatBC can be co-purified (Bageshwar *et al.*, 2009, Tarry *et al.*, 2009). Other studies have found the signal peptide and the early mature domain binding to the TatBC complex in an inverted hairpin conformation where the h-domain is the same height as the TatB TMD, potentially similar to deep-insertion mode (Frobel *et al.*, 2012, Hou *et al.*, 2006).

1.7.5 Tat translocation

Four steps have been defined in Tat translocation, shown schematically in Figure 1.17. Firstly, the signal peptide is recognised by the TatBC complex. This is followed by the assembly of the final translocase with the polymerisation of TatA. Subsequently the passenger domain is translocated into the periplasm where the signal peptide is cleaved and the translocase dissociates.

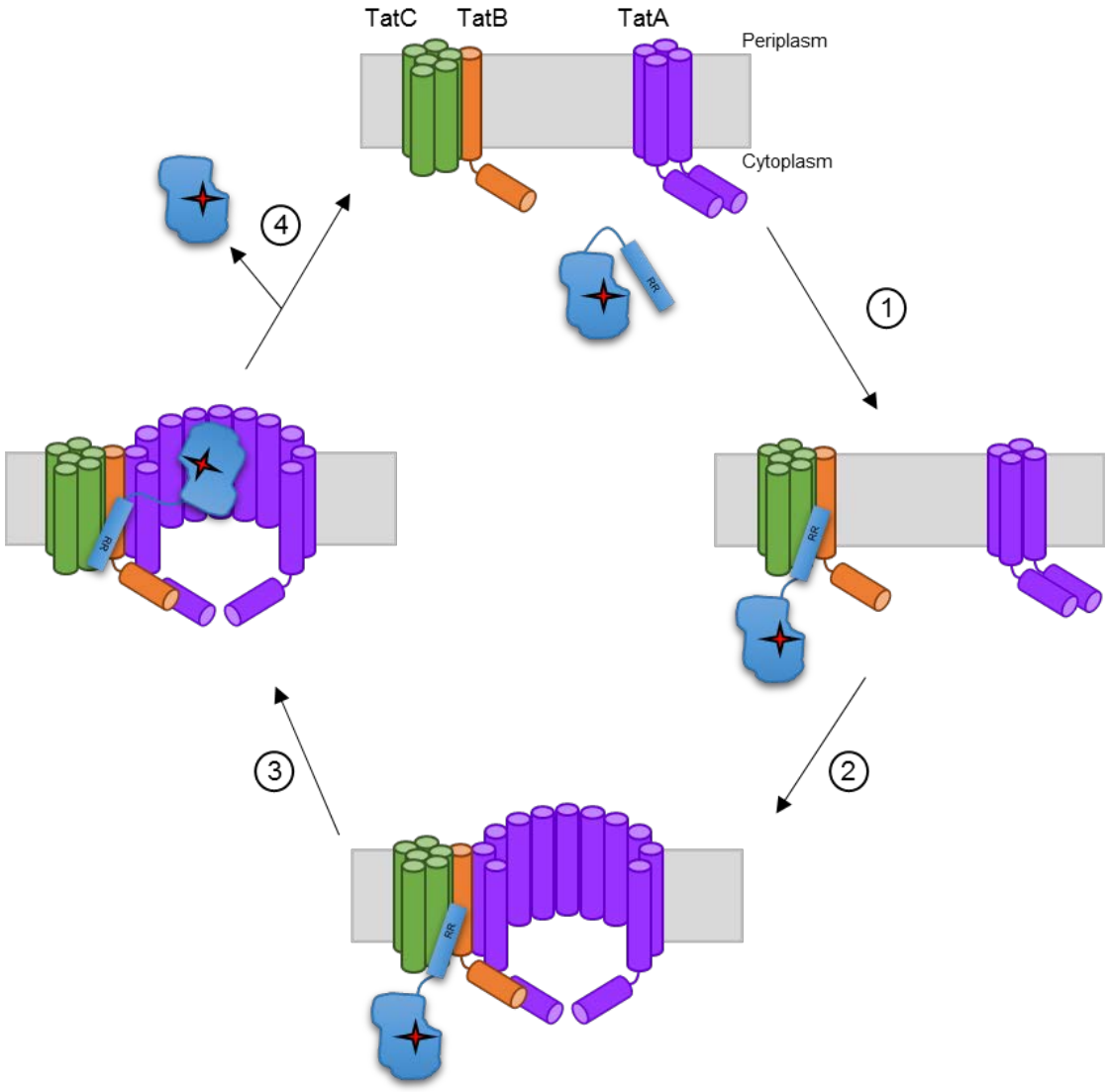


Figure 1.17 Schematic model for the Tat translocation cycle in *E. coli*.
The Tat translocase is made up of three main proteins; TatA (purple), TatB (orange) and TatC (green). Under resting conditions these proteins are found in two distinct complexes, TatBC with 1:1 stoichiometry and TatA in tetrameric form (Palmer & Berks, 2012, Leake *et al.*, 2008). 1) The translocation cycle is initiated by the interaction of the TatBC complex and the signal peptide of a folded Tat substrate. 2) Upon substrate binding to TatBC, TatA polymerisation occurs forming a pore across the cytoplasmic membrane. 3) The substrate then crosses the membrane through the TatA channel. 4) Following translocation the signal peptide is cleaved by a signal peptidase. TatA subunits dissociate from TatBC and depolymerise back to tetramers.

Initially the signal peptide of the Tat substrate binds to the TatBC complex (Cline & Mori, 2001, Gerard & Cline, 2007). This binding is independent of TatA and does not require the PMF (Gerard & Cline, 2007). The substrate might be targeted to the translocation machinery by chaperones or through 2D diffusion along the membrane surface (Papish *et al.*, 2003, Shanmugham *et al.*, 2006, Bageshwar *et al.*, 2009).

The formation of a TatBC-substrate complex triggers association of TatA with the TatBC complex, leading to the formation of an active translocon (Cline & McCaffery, 2007, Mori & Cline, 2002, Maurer *et al.*, 2010). This is a transient interaction of TatA with TatBC which is dependent upon the PMF. Chemical crosslinking has shown that thylakoid TatA undergoes polymerisation which is induced by the TatBC-substrate complex and the PMF (Dabney-Smith & Cline, 2009, Dabney-Smith *et al.*, 2006). *In vivo* studies with *E. coli* TatA labelled with YFP, have seen that low order oligomers were formed without TatBC but when TatBC was present TatA-YFP formed larger complexes (Leake *et al.*, 2008). However, crosslinking and co-purification experiments have found monomeric TatA associated with TatBC before substrate binding, potentially as a nucleation site for further TatA polymerisation (Frobel *et al.*, 2011, Taubert *et al.*, 2015, de Leeuw *et al.*, 2002). The folding state of the substrate appears to be monitored by polymerised TatA which is indicated by an *in vitro* transport reporter protein (PhoA) fused to the signal peptide of TorA that stalls with the PhoA passenger domain as a protease resistant fragment interacting with TatA (Panahandeh *et al.*, 2008).

There are two potential models for the translocation system, the first proposes a form-fitting channel lined by TatA, occurring in varying diameters to enable translocation of differently sized substrates (Sargent *et al.*, 2006, Gohlke *et al.*, 2005). Alternatively, a second model proposes that the bilayer is weakened and thinned by oligomerisation of TatA and this instability is harnessed to generate transient pores for translocation (Dabney-Smith & Cline, 2009, Bruser & Sanders, 2003), this model is supported by modelling analysis (Rodriguez *et al.*, 2013).

Following translocation the mature substrate is released *trans* side of the membrane and the signal peptide can be cleaved in the c-region at the AxA site by signal peptidase (Luke *et al.*, 2009, Yahr & Wickner, 2001). The translocation machinery dissociates upon substrate release which is shown by a loss of crosslinking between TatA and TatBC (Mori & Cline, 2002).

1.8 Respiratory chain

The aerobic electron transport chain (respiratory chain) of mitochondria and many bacteria has four complexes which shuttle electrons to generate a transmembrane electrochemical proton gradient (Crofts, 2004). The transfer of electrons is coupled to the movement of protons from a region of low proton concentration and negative electrical potential (N phase) to a region of high proton concentration and positive electrical potential (P phase). This generates an electrochemical proton gradient across the membrane which can be used in the cell as an energy source, particularly for the synthesis of ATP.

In mitochondria the components of the respiratory chain are known as Complex I to IV. Complex I is a NADH-ubiquinone reductase (NADH dehydrogenase) which catalyses the transfer of electrons from the electron donor, NADH, to ubiquinone (UQ). Complex II, succinate dehydrogenase, also catalyses the transfer of electrons to UQ but from the electron donor reduced FAD. Complex III, ubiquinol-cytochrome *c* oxidoreductase, oxidises UQH_2, transferring two electrons to cytochrome *c*, in a reaction known as the Q cycle. Following this, a mobile cytochrome *c* (known as cytochrome c_2 in bacteria) shuttles the electrons to Complex IV (cytochrome *c* oxidase) which transfers them to a final acceptor, O_2.

1.8.1 Cytochrome bc_1 complex and the Rieske protein

The central component of this system is Complex III, the ubiquinol-cytochrome *c* oxidoreductase which in bacteria and mitochondria is known as the cytochrome bc_1 complex. It has a homologue in plant and algal chloroplasts and in cyanobacteria is known as the cytochrome b_6f complex (Darrouzet *et al.*, 2004, Berry *et al.*, 2000). This complex has three conserved redox active components: cytochrome *b*, cytochrome c_1 or *f* and an FeS cluster containing protein (Berry *et al.*, 2000). It functions to catalyse an

electron transfer reaction from a lipophilic substrate ubiquinol (UQH_2) to cytochrome *c* coupled with proton translocation across the membrane.

Cytochrome *b* consists of 8TMDs and co-ordinates two *b*-type hemes; b_L which is located on the periplasmic side (P phase) and is a low potential heme and b_H which is located on the cytoplasmic side (N phase) and is a high potential heme. There are also two possible binding sites for the quinone moieties adjacent to either heme: Q_o and Q_i. Cytochrome c_1 is membrane-bound by a C-terminal anchor and has a periplasmic domain containing a *c* type heme. In comparison, the FeS protein (known as Rieske) has an N-terminal membrane anchor connected to a globular periplasmic domain containing an FeS cluster, [2Fe-2S]. The flexible linker of these domains has the ability to rotate allowing the catalytic domain to rearrange its position. The functional form of this complex is as a dimer where the Rieske protein interacts with both monomers, its catalytic domain interacting with one and the TMD with the other (Xia *et al.*, 1997, Kurisu *et al.*, 2003). The mobile cytochrome c_2 binds cytochrome c_1 substoichiometrically as it only binds to one monomer in the dimer.

The Q cycle involves a two electron carrier, ubiquinol (UQH_2), to bind to cytochrome *b* where it can interact with a one electron carrier, Rieske, permitting the transfer of electrons to the soluble cytochrome *c*, as modelled in Figure 1.18 (Cooley, 2013). UQH_2 binds to the Q_o site of cytochrome *b*, one electron is transferred to the FeS domain of Rieske and follows a high potential electron transfer chain. There is a 14 Å limit for transfer of electrons, therefore, it is vital that the catalytic domain FeS is repositioned to dock onto cytochrome c_1 for electron transfer, this involves movement of 20 Å (Iwata *et al.*, 1998, Zhang *et al.*, 1998, Crofts, 2004). Following the transfer of this electron to the membrane bound *c*-type heme of cytochrome c_1, which passes it on to a mobile electron carrier protein cytochrome c_2, this protein is able to shuttle the electron to the cytochrome *c* oxidase complex. This high potential reaction is coupled to the transfer of two protons across the membrane. The second electron is passed to the b_L heme of cytochrome *b* where it is transferred to the b_H heme and onto the quinone in the Q_i site. This is a low

potential chain of electron transfer. This entire process is repeated resulting in a net transfer of two electrons to cytochrome *c* and four protons across the membrane.

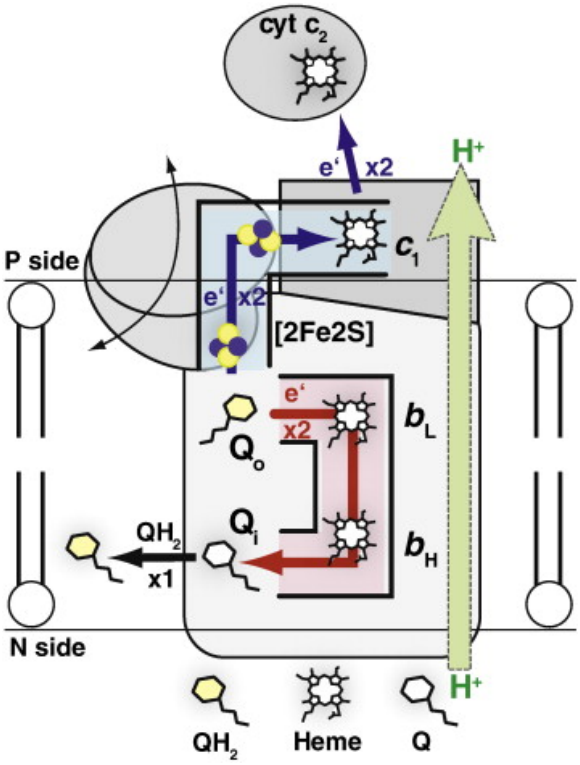


Figure 1.18 Cartoon model of the turnover of the cytochrome *bc*$_1$ complex.
QH_2 binds to the Q_o site on Cytochrome *b* and its two electrons are transferred through either a high (blue) or low (red) potential electron transport chains. The high potential chain involves transfer through the Rieske FeS protein to cytochrome c_1 and onto the mobile electron carrier cytochrome c_2. The low potential chain involves electron cycling through the *b*-type hemes on cytochrome *b*. N side is cytoplasmic whereas P side is periplasmic. Taken from Cooley (2013).

The biogenesis of cytochrome *bc*$_1$ involves both the Sec and the Tat translocation pathways. Apo-cytochrome c_1 and apo-cytochrome *b* are both inserted into the membrane by the Sec pathway (Thony-Meyer & Kunzler, 1997). The *c*-type heme is covalently inserted into cytochrome c_1 by the system I cytochrome *c* biogenesis machinery, whereas, the *b*-type heme is spontaneously inserted into cytochrome *b*. The cofactor insertion into Rieske is coordinated by Cys residues which form a disulphide bond for stability (Iwata *et al.*, 1998, Hunte *et al.*, 2000). The domain is folded as a globular structure which incorporates the FeS cluster close to the surface and is connected to the N-terminal membrane anchor through a flexible tether linker. Failure to insert the FeS cluster into Rieske impairs the assembly of this protein into the

cytochrome bc_1 complex (Gutierrez-Cirlos *et al.*, 2002). In the *B. subtilis* Rieske protein there is a sequential and hierarchical quality control for cofactor binding and disulphide bond formation to prevent membrane insertion of unfolded, cofactor-less proteins (Goosens *et al.*, 2014).

In contrast to the cytochromes, the Rieske protein has been found to be a Tat substrate in bacteria and plants, for example in *Legionella pneumophila*, *Paracoccus dentrificans* and in plant thylakoids (De Buck *et al.*, 2007, Molik *et al.*, 2001, Bachmann *et al.*, 2006). There are two classes of Tat-dependent substrates that are inserted into the membrane; C-terminal anchored and N-terminal anchored (Berks, 1996). The Rieske protein is in the latter class, the N-terminal anchor is composed of the Tat targeting signal sequence which is uncleaved and remains in the membrane (Aldridge *et al.*, 2008, Bachmann *et al.*, 2006, De Buck *et al.*, 2007, Molik *et al.*, 2001, Meloni *et al.*, 2003).

1.8.2 Actinobacterial cytochrome bc_1 complex

In Actinobacteria, the major route for aerobic quinol oxidation is through the cytochrome bc_1 and cytochrome aa_3 complexes. Studies in *Corynebacterium glutamicum* have identified the genes coding for cytochrome bc_1 and cytochrome aa_3 complexes (Niebisch & Bott, 2001, Sone *et al.*, 2001). Cytochrome bc_1 genes are *qcrA*, *qcrB* and *qcrC* and these are located in a cluster, downstream of subunit III of cytochrome aa_3, *ctaE*. The other cytochrome aa_3 subunit genes are located elsewhere, *ctaD* and *ctaC.* These genes are homologous in *Mycobacteria* and *Streptomyces* (Sone *et al.*, 2001).

There are distinct differences between the cytochrome bc_1 complex in Actinobacteria, compared to other bacteria (Niebisch & Bott, 2001, Sone *et al.*, 2001). The cytochrome *b* (QcrB) has 9 TMD instead of 8 and a small globular domain is present at the *trans* side of the membrane. The cytochrome c_1 (QcrC) has been found to have two CxxCH motifs which are important for covalent heme attachment. This implies that it is a di-heme *c*-type cytochrome. As there are no soluble *c*-type cytochromes present in these bacteria it is predicted that this additional *c*-type heme functions as the cytochrome c_2 and shuttles

the electrons from cytochrome c_1 to cytochrome aa_3 complex. This functionality would only be possible if the two complexes were interacting. Both complexes can be co-purified together indicating the potential for a supercomplex of cytochrome bc_1-aa_3 (Niebisch & Bott, 2003). The Rieske protein (QcrA) is also different in these cytochrome bc_1 complexes as it has increased hydrophobicity with 3 TMDs N-terminal to the globular domain instead of 1 TMD. Despite these differences in the complexes involved in proton transfer and the importance of these bacteria as agents of disease, not much is known about the essential respiratory chain in these organisms.

1.9 Co-operation between Sec and Tat translocation

The protein sequence alignment seen in Figure 1.19 shows that Actinobacterial Rieske proteins (from *S. coelicolor*, *C. glutamicum* and *M. tuberculosis*) have three TMDs directly before the catalytic FeS globular domain, this is in comparison to other bacterial Rieske proteins. The increased TMDs causes a biosynthetic problem for insertion of Rieske proteins as they are usually Tat-dependent and it is not known whether the Tat machinery is able to insert more than one TMD. From the alignment of amino acid sequences of Rieske proteins, the essential Tat motif is located N-terminal of TMD3, suggesting that the Tat machinery is involved in the insertion of TMD3 and the translocation of the cofactor domain. Keller *et al.* (2012) investigated the insertion of *S. coelicolor* Rieske within the heterologous host *E. coli*. Using reporters fused to the transmembrane component of Rieske they discovered that the Tat motif is recognised and TMD3 and the FeS domain are translocated by the Tat machinery. In comparison, the first two TMD interact with the Sec machinery for insertion, probably through the co-translational SRP pathway and has been shown to be YidC-dependent.

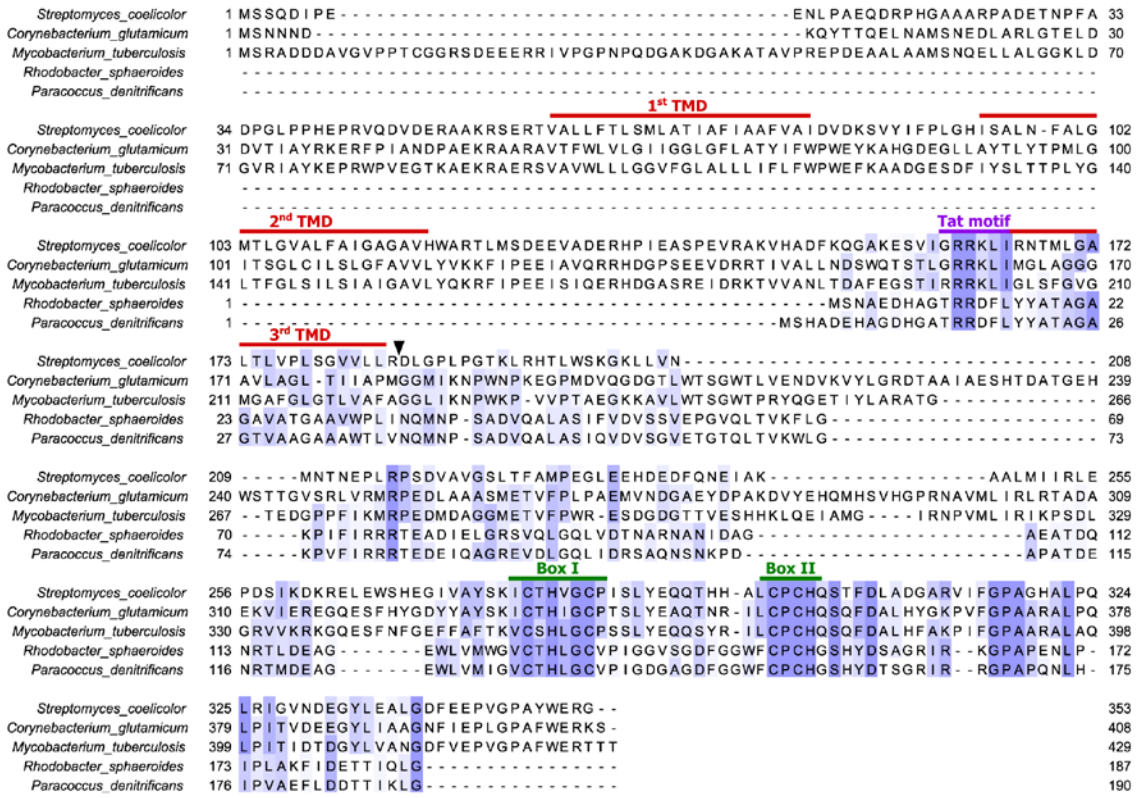


Figure 1.19 Protein sequence alignment of the Rieske iron-sulfur proteins from different bacteria. Proteins were aligned using Clustal WS and Jalview (Waterhouse *et al.*, 2009). The longer Rieske FeS proteins of actinobacterial representatives (*Streptomyces coelicolor*, *Corynebacterium glutamicum* and *M. tuberculosis*) are aligned alongside the well-characterised Rieske proteins of *Rhodobacter sphaeroides* and *Paracoccus dentrificans*. Transmembrane domains (TMD) were predicted using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and are marked in red above the alignment. The consensus twin-arginine (Tat) motif is highlighted in purple. Conserved boxes I and II that coordinate the 2Fe-2S cluster are highlighted in green. The arrow indicates the position after which the reporter proteins were fused. The differences in shading (from light blue to dark) refer to the level of amino acid conservation between the different species.

From this data, a model of insertion was suggested (Figure 1.20), where TMD1 and TMD2 are inserted into the cytoplasmic membrane by the Sec machinery with the assistance of YidC. Following the release of the polypeptide by the Sec machinery, the FeS cluster is assembled into the globular domain in the cytoplasm before the Tat machinery takes over and inserts the final TMD and translocates the cofactor containing domain into the periplasm.

These results were confirmed in *S. coelicolor*, the native organism, where the Tat system was demonstrated to be essential for the activity of the cytochrome bc_1 complex. This is due to the requirement for the Tat machinery to translocate the FeS domain of Rieske. However, the Rieske protein is stably inserted into the membrane in the absence of the

Tat system implying that the Rieske protein is dual targeted to the membrane within *S. coelicolor* (Hopkins *et al.*, 2014). This is the first protein known to require co-operation of the Sec and Tat pathways for its insertion.

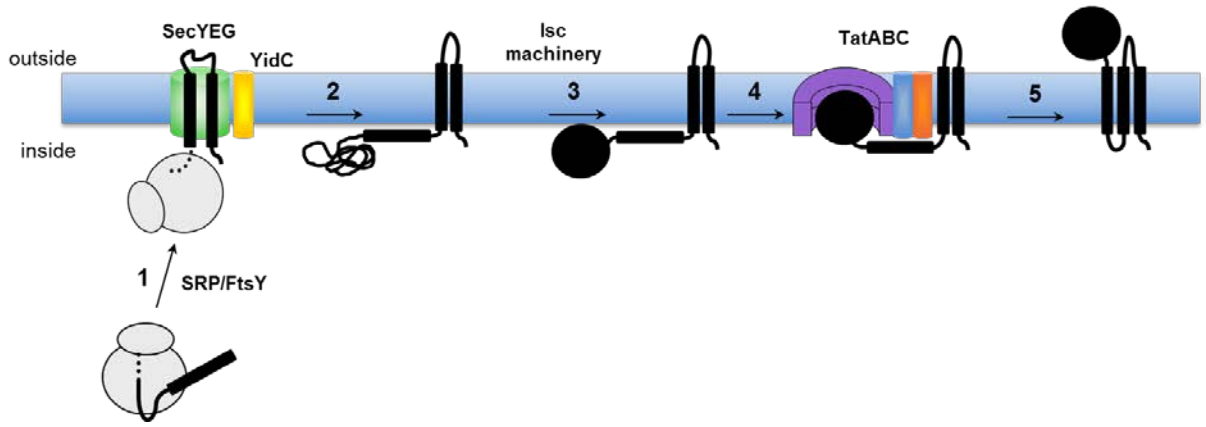


Figure 1.20 Model for the biogenesis of the *S. coelicolor* Rieske protein.
1) SRP binds to the first hydrophobic segment of Rieske emerging from the ribosome and guiding the complex to the Sec machinery. 2) The first two TMDs of Rieske are inserted into the membrane co-translationally. 3) The FeS cluster is inserted into Rieske, leading to the folding of the globular domain. 4) The Tat system inserts TMD3 and translocates the FeS cluster-containing domain into the periplasm. 5) Fully assembled Rieske is released into the membrane, where it interacts with partner proteins in the cytochrome bc_1 complex. Adapted from Keller *et al.* (2012).

1.10 Aims

The overall aim of this study is to define key regions of the *S. coelicolor* Rieske protein that are important for the interaction with the Sec and Tat machineries during its dual targeted insertion. Specifically, to identify which features of the Rieske protein promote the protein's release from the Sec pathway and which facilitate recognition of an internal signal sequence by the Tat machinery. This work used reporter proteins fused to the TMDs of *S. coelicolor* Rieske and expressed in the heterologous host *E. coli*.

First, it was essential to investigate potential fusion reporters to ensure robust and reliable assays that indicated clear positive and negative results (Chapter 3). Following this, a full programme of mutagenesis was undertaken to elucidate any important features for interactions with the Sec or Tat machinery, which was then followed by biochemical approaches to confirm these results (Chapters 4 and 5).

2 Materials and Methods

2.1 Bacterial strains

All strains used in this study are derived from *Escherichia coli* K-12 and are listed in Table 2.1. Strain DH5α was used for general transformation of plasmid vectors and cloning. Strains MC4100 and DADE were used for all work with β-lactamase-encoding constructs and MCDSSAC and MCDSSACΔtat were used for all work with AmiA-encoding constructs, as appropriate *tat*$^+$ and *tat*$^-$ strains, respectively.

Table 2.1 Bacterial strains used in this study.

Strains	Genotype	Resistance	Reference
MC4100	F$^-$Δ*lac*U169 *araD139 rpsL150 relA1 ptsF rbs flbB5301*	None	(Casadaban & Cohen, 1979)
DADE	MC4100 Δ*tatABCD* Δ*tatE*	None	(Wexler *et al.*, 2000)
MCDSSAC	MC4100 *amiA*Δ2-33, *amiC*Δ2-32	None	(Ize *et al.*, 2003)
MCDSSACΔtat	MCDSSAC, Δ*tatABC*::Apra	Apra	(Keller *et al.*, 2012)
DH5α	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K^- m_K^+), λ–*	None	(Grant *et al.*, 1990)

2.2 Materials

2.2.1 Growth media and growth conditions

Growth media and additives used in this study are outlined in Table 2.2, Table 2.3 and Table 2.4. Strains were commonly grown aerobically overnight at 37°C in Luria-Bertani (LB) liquid broth with vigorous shaking at 200 rpm. Growth on solid media was also performed at 37°C.

Long term storage of strains was carried out at -80°C by the addition of a final concentration of 25% glycerol to a stationary phase culture, with this being flash frozen in liquid nitrogen before storage.

Table 2.2 Growth media used in this study.

Media name	Component	Concentration
Luria-Bertani (LB) medium	Tryptone	1.0 % (w/v)
	Yeast extract	0.5 % (w/v)
	NaCl	1.0 % (w/v)
LB agar	Agar	1.5 % (w/v)
	Tryptone	1.0 % (w/v)
	Yeast extract	0.5 % (w/v)
	NaCl	1.0 % (w/v)
Low salt LB agar	Agar	1.5 % (w/v)
	Tryptone	1.0 % (w/v)
	Yeast extract	0.5 % (w/v)
	NaCl	0.5 % (w/v)
TSB/DMSO medium (transformation buffer)	Tryptone	1.0 % (w/v)
	Yeast extract	0.5 % (w/v)
	NaCl	1.0 % (w/v)
	DMSO	5.0 % (v/v)
	$MgSO_4$	50 mM
	Polyethylene glycol (MW 6000)	10.0 % (w/v)

Table 2.3 Additives used throughout this study.

Supplement	Stock solution	Final concentration
Sodium dodecyl sulphate (SDS)	20 % (w/v)	1-2 %

Table 2.4 Antibiotics used in this study with their stock and working conditions.
Stock solutions of amp, Kan and apra were prepared in distilled water. Cm was prepared in 80 % (v/v) ethanol. All stock solutions were filter sterilised prior to use.

Antibiotic	Stock concentration	Final concentration
Ampicillin (amp)	125 mg/ml	125 µg/ml
Kanamycin (Kan)	50 mg/ml	50 µg/ml
Apramycin (apra)	25 mg/ml	25 µg/ml
Chloramphenicol (Cm)	25 mg/ml	25 µg/ml

2.3 Buffers and solutions

Table 2.5 General buffers and solutions used in this study

Buffer/solution	Composition	Concentration
APS	Ammonium persulphate	10 % (w/v)
DNA loading dye	Bromophenol blue	0.25 % (w/v)
	Xylene cyanol blue	0.25 % (w/v)
	Sucrose	40 % (w/v)
6x sample buffer	Tris-HCl (pH 6.8)	62.5 mM
	SDS	2 % (w/v)
	β-mercaptoethanol	15 % (v/v)
	Glycerol	25 % (v/v)
	Bromophenol blue	0.01 % (w/v)
Laemmli sample buffer (2x)	Tris-HCl (pH 6.8)	65.8 mM
	SDS	2.1 % (w/v)
	β-mercaptoethanol	355 mM
	Glycerol	26.3 % (w/v)
	Bromophenol blue	0.01 % (w/v)
HEPES/$MgCl_2$ (pH 6.8)	HEPES	50 mM
	$MgCl_2$	5 mM
SDS running buffer (10x)	Tris-HCl (pH 8.3)	250 mM
	Glycine	1.92 M
	SDS	1.0 % (w/v)
TAE buffer	Tris-HCl (pH 8.0)	40 mM
	Glacial acetic acid	1.142 % (v/v)
	EDTA	1 mM
Tris Buffered Saline (TBS)	Tris-HCl (pH 7.5)	20 mM
	NaCl	137 mM
TBS/Tween (TBST)	Tris-HCl (pH7.5)	20 mM
	NaCl	137 mM
	Tween®20	0.1 % (v/v)
Tris-glycine transfer buffer	Tris/HCl (pH8.8)	25 mM
	Glycine	192 mM
	Methanol	10 % (v/v)
Buffer 1 (membrane buffer)	Tris-HCl (pH 7.5)	20 mM
	NaCl	200 mM
Buffer 2 (membrane resuspension buffer)	Tris-HCl (pH7.5)	50 mM
	$MgCl_2$	5 mM
	Glycerol	10 % (v/v)

2.4 Molecular biology techniques

2.4.1 Preparation of competent cells and transformation with plasmid DNA

50 ml of LB broth (with required antibiotics) was inoculated at 1:100 dilution from a stationary phase culture and grown aerobically at 37°C with shaking at 200 rpm until an OD_{600} of 0.4 was achieved. After incubation for 20 min on ice, cells were harvested by centrifugation at 3000 x *g* for 10 min. Cell pellets were resuspended in ice cold TSB/DMSO medium and kept on ice for a minimum of 10 min before use or frozen in liquid nitrogen and stored at -80°C. Transformation was performed by incubating 100 µl of competent cells with 1 µl of plasmid (or 10 µl of ligation mix) for 15 min on ice. Cells were subjected to heat-shock at 42°C for 90 s followed by an additional 2 min incubation on ice. 1 ml of LB was added and cells were grown, with shaking, at 37°C for 1 h. Cells were then pelleted by centrifugation at 16000 x *g* for 1 min. This was followed by plating of cells onto LB agar plates containing required antibiotics and incubation overnight at 37°C.

2.4.2 Plasmid DNA preparation

Plasmid extraction from *E. coli* strains was undertaken using the QIAprep® Spin Miniprep kit (Qiagen), as per the manufacturer's instructions, this procedure uses an alkaline lysis method developed by Birnboim and Doly (Birnboim & Doly, 1979). A single colony was used to inoculate a 5 ml culture which was grown until stationary phase before centrifugation for 10 min at 3000 x *g*. Cells were subsequently resuspended in lysis buffer and then treated with neutralisation buffer before isolation of the plasmid DNA through adsorption onto a silica membrane under high salt conditions. After washing, the DNA was eluted from the membrane in 30 µl of either a low-salt elution buffer or ultra-pure water.

The plasmids used in this study are shown in Table 2.6.

Table 2.6 Bacterial plasmids used in this study.

Plasmid name	Vector	Characteristics	Antibiotic resistance	Reference
pSUPROM	pSU40	Expression vector with constitutive *E. coli tatA* promoter	Kan	(Jack *et al*., 2004)
pSU18	pSU18	Expression vector with constitutive *E. coli lac* promoter	Cm	(Bartolome *et al.,* 1991)
pSUPROM Rieske-Bla	pSUPROM	pSUPROM producing aa 1-185 of Sco2149 fused to aa 24-286 of β-lactamase obtained by PCR amplification from pBR322 (source of Bla from transposon Tn3)	Kan	This study
pSUPROM Rieske-AmiA	pSUPROM	pSUPROM producing aa 1-185 of Sco2149 fused to aa 32-289 of *E. coli* AmiA	Kan	This study
pSU18 AmiA	pSU18	pSU18 producing aa 1-289 of *E. coli* AmiA	Cm	(Ize *et al*., 2003)
pSUPROM AmiA	pSUPROM	pSUPROM producing aa 1-289 of *E. coli* AmiA	Kan	This study
pSUPROM Bla	pSUPROM	pSUPROM producing aa 24-286 of β-lactamase obtained by PCR amplification from pBR322	Kan	This study
pSU18 Rieske-AmiA	pSU18	pSU18 producing aa 1-185 of Sco2149 fused to aa 32-289 of *E. coli* AmiA	Cm	(Keller *et al*., 2012)
pSU18 RRKK-Rieske-AmiA	pSU18	pSU-TM123-AmiA with substitution of aa 161-162 of Sco2149 from RR to KK	Cm	(Keller *et al*., 2012)
pHASoxYZ	pSU20	pFLUOR2A_YFP_C producing *P. panthotrophus soxYZ* under the control of *E. coli tatA* promoter with the signal peptide of SoxY replaced by an N-terminal haemagglutinin tag	Cm	(Bauer *et al*., 2011)
pSUPROM MTB Rieske-AmiA	pSUPROM	pSUPROM producing aa 1-227 of *M. tuberculosis* H37Rv QcrA codon optimised for *E. coli* fused to aa 32-289 of *E. coli* AmiA	Kan	This study
pSUPROM MTB Rieske-Bla	pSUPROM	pSUPROM producing aa 1-227 of *M. tuberculosis* H37Rv QcrA codon optimised for *E. coli* fused to aa 24-286 of β-lactamase obtained by PCR amplification from pBR322	Kan	This study
pSUPROM Rieske-Bla RHHR	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 133-134 of Rieske from RH to HR	Kan	This study

pSUPROM Rieske-Bla RHKK	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 133-134 of Rieske from RH to KK	Kan	This study
pSUPROM Rieske-Bla A138P	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 138 of Rieske from A to P	Kan	This study
pSUPROM Rieske-Bla A144P	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 144 of Rieske from A to P	Kan	This study
pSUPROM Rieske-Bla A148P	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 148 of Rieske from A to P	Kan	This study
pSUPROM Rieske-Bla A154P	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 154 of Rieske from A to P	Kan	This study
pSUPROM Rieske-Bla M124A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 124 of Rieske from M to A	Kan	This study
pSUPROM Rieske-Bla S125A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 125 of Rieske from S to A	Kan	This study
pSUPROM Rieske-Bla D126A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 126 of Rieske from D to A	Kan	This study
pSUPROM Rieske-Bla E127A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 127 of Rieske from E to A	Kan	This study
pSUPROM Rieske-Bla M124L	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 124 of Rieske from M to L	Kan	This study
pSUPROM Rieske-Bla S125L	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 125 of Rieske from S to L	Kan	This study
pSUPROM Rieske-Bla D126L	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 126 of Rieske from D to L	Kan	This study
pSUPROM Rieske-Bla E127L	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 127 of Rieske from E to L	Kan	This study
pSUPROM Rieske-Bla RRKK	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 161-162 of Rieske from RR (Tat motif) to KK	Kan	This study
pSUPROM Rieske-Bla RRKQ	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 161-162 of Rieske from RR (Tat motif) to KQ	Kan	This study

pSUPROM Rieske-Bla RRAA	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 161-162 of Rieske from RR (Tat motif) to AA	Kan	This study
pSUPROM Rieske-Bla RRAD	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 161-162 of Rieske from RR (Tat motif) to AD	Kan	This study
pSUPROM Rieske-Bla ΔRR	pSUPROM	pSUPROM Rieske-Bla with deletion of aa (RR Tat motif) 161-162 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-122	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-122 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ123-127	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 123-127 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ128-132	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 128-132 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ133-137	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 133-137 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ138-142	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 138-142 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ143-147	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 143-147 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ148-152	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 148-152 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ153-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 153-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-127	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-127 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ128-137	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 128-137 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ138-147	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 138-147 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ148-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 148-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ138-152	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 138-152 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ123-137	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 123-137 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ128-142	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 128-142 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ133-147	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 133-147 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-132	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-132 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ143-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 143-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-137	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-137 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ138-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 138-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-142	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-142 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ123-147	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 123-147 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ128-152	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 128-152 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ133-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 133-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-147	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-147 of Rieske	Kan	This study

pSUPROM Rieske-Bla Δ123-152	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 123-152 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ128-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 128-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ123-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 123-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-152	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-152 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-153	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-153 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-154	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-154 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-155	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-155 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-156	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-156 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ118-157	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-157 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ126-127	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 126-127of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ127-128	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 127-128 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ126-128	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 126-128 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ131-132	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 131-132 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ137Δ141	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 137 & 141of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ149Δ156	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 149 & 156 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ131-132 Δ141	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 131 & 132 & 141 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ126-128 Δ131-132	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 126-128 &131-132 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ126-128 Δ137 Δ141	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 126-128 &137 &141 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ131-132 Δ137 Δ141	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 131-132 & 137 &141 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ126-128 Δ131-132 Δ137 Δ141	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 126-128 &131-132 &137 & 141 of Rieske	Kan	This study
pSUPROM Rieske-Bla Δ126-128 Δ131-132 Δ137 Δ141 Δ149 Δ156	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 126-128 &131-132 &137 & 141 &149 & 156 of Rieske	Kan	This study
pSUPROM Rieske-Bla D131A E132A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 131 & 132 of Rieske to A	Kan	This study
pSUPROM Rieske-Bla E137A E141A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 137 & 141 of Rieske to A	Kan	This study

pSUPROM Rieske-Bla D126A E127A E128A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 126 & 127 & 128 of Rieske to A	Kan	This study
pSUPROM Rieske-Bla D131K E132K	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 131 & 132 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla E137K E141K	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 137 & 141 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla D126K E127K E128K	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 126 & 127 & 128 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla INS D129 E130 E131	pSUPROM	pSUPROM Rieske-Bla with insertion of DEE at positions aa 129-131 of Rieske	Kan	This study
pSUPROM Rieske-Bla D126K E127K E128K E137K E141K	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 126-128 & 137 & 141 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla D126K E127K E128K D131K E132K E137K E141K	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 126-128 & 131-132 & 137 & 141 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla R185C	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 185 of Rieske to C	Kan	This study
pSUPROM Rieske-Bla R185A	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 185 of Rieske to A	Kan	This study
pSUPROM Rieske-Bla P177L S179L G180L	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 177 &179 & 180 of Rieske to L	Kan	This study
pSUPROM Rieske-Bla S179L G180L	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 179 & 180 of Rieske to L	Kan	This study
pSUPROM Rieske-Bla S179L		pSUPROM Rieske-Bla with substitution of aa 179 of Rieske to L	Kan	This study
pSUPROM Rieske-Bla G180L	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 180 of Rieske to L	Kan	This study
pSUPROM Rieske-Bla V158K	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 158 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla ∆118-153 V158K	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-153 of Rieske & substitution of aa 158 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla ∆118-154 V158K	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-154 of Rieske & substitution of aa 158 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla ∆118-155 V158K	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-155 of Rieske & substitution of aa 158 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla ∆118-156 V158K	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-156 of Rieske & substitution of aa 158 of Rieske to K	Kan	This study

pSUPROM Rieske-Bla Δ118-157 V158K	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-157 of Rieske & substitution of aa 158 of Rieske to K	Kan	This study
pSUPROM Rieske-Bla V158C	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 158 of Rieske to C	Kan	This study
pSUPROM Rieske-Bla Δ118-153 V158C	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-153 of Rieske & substitution of aa 158 of Rieske to C	Kan	This study
pSUPROM Rieske-Bla Δ118-154 V158C	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-154 of Rieske & substitution of aa 158 of Rieske to C	Kan	This study
pSUPROM Rieske-Bla Δ118-155 V158C	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-155 of Rieske & substitution of aa 158 of Rieske to C	Kan	This study
pSUPROM Rieske-Bla Δ118-156 V158C	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-156 of Rieske & substitution of aa 158 of Rieske to C	Kan	This study
pSUPROM Rieske-Bla Δ118-157 V158C	pSUPROM	pSUPROM Rieske-Bla with deletion of aa 118-157 of Rieske & substitution of aa 158 of Rieske to C	Kan	This study
pSUPROM Rieske-Bla V82C	pSUPROM	pSUPROM Rieske-Bla with substitution of aa 82 of Rieske to C	Kan	This study
pSUPROM Rieske-AmiA RHHR	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 133-134 of Rieske from RH to HR	Kan	This study
pSUPROM Rieske-AmiA RHKK	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 133-134 of Rieske from RH to KK	Kan	This study
pSUPROM Rieske-AmiA A138P	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 138 of Rieske from A to P	Kan	This study
pSUPROM Rieske-AmiA A144P	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 144 of Rieske from A to P	Kan	This study
pSUPROM Rieske-AmiA A148P	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 148 of Rieske from A to P	Kan	This study
pSUPROM Rieske-AmiA A154P	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 154 of Rieske from A to P	Kan	This study
pSUPROM Rieske-AmiA M124A	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 124 of Rieske from M to A	Kan	This study
pSUPROM Rieske-AmiA S125A	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 125 of Rieske from S to A	Kan	This study

pSUPROM Rieske-AmiA D126A	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 126 of Rieske from D to A	Kan	This study
pSUPROM Rieske-AmiA E127A	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 127 of Rieske from E to A	Kan	This study
pSUPROM Rieske-AmiA M124L	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 124 of Rieske from M to L	Kan	This study
pSUPROM Rieske-AmiA S125L	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 125 of Rieske from S to L	Kan	This study
pSUPROM Rieske-AmiA D126L	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 126 of Rieske from D to L	Kan	This study
pSUPROM Rieske-AmiA E127L	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 127 of Rieske from E to L	Kan	This study
pSUPROM Rieske-AmiA RRKK	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 161-162 of Rieske from RR (Tat motif) to KK	Kan	This study
pSUPROM Rieske-AmiA RRKQ	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 161-162 of Rieske from RR (Tat motif) to KQ	Kan	This study
pSUPROM Rieske-AmiA RRAA	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 161-162 of Rieske from RR (Tat motif) to AA	Kan	This study
pSUPROM Rieske-AmiA RRAD	pSUPROM	pSUPROM Rieske-AmiA with substitution of aa 161-162 of Rieske from RR (Tat motif) to AD	Kan	This study
pSUPROM Rieske-AmiA ΔRR	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa (RR Tat motif) 161-162 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-122	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-122 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ123-127	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 123-127 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ128-132	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 128-132 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ133-137	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 133-137 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ138-142	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 138-142 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ143-147	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 143-147 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ148-152	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 148-152 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ153-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 153-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ158-162	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 158-162 of Rieske	Kan	This study

pSUPROM Rieske-AmiA Δ118-127	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-127 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ128-137	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 128-137 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ138-147	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 138-147 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ148-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 148-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-132	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-132 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ123-137	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 123-137 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ128-142	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 128-142 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ133-147	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 133-147 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ138-152	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 138-152 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ143-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 143-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-137	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-137 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ138-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 138-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-142	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-142 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ123-147	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 123-147 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ128-152	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 128-152 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ133-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 133-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-147	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-147 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ123-152	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 123-152 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ128-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 128-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ123-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 123-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-152	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-152 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-153	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-153 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-154	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-154 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-155	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-155 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-156	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-156 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ118-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 118-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ124-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 124-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ125-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 125-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ126-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 126-157 of Rieske	Kan	This study
pSUPROM Rieske-AmiA Δ127-157	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 127-153 of Rieske	Kan	This study
pSUPROM Rieske-AmiA HWAKES	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 121-154 of Rieske	Kan	This study

pSUPROM Rieske-AmiA HWKES	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 120-154 of Rieske	Kan	This study
pSUPROM Rieske-AmiA HWES	pSUPROM	pSUPROM Rieske-AmiA with deletion of aa 120-155 of Rieske	Kan	This study
pSUPROM MTB Rieske-Bla RRAD	pSUPROM	pSUPROM MTB Rieske-Bla substitution of aa 199 & 200 of Rieske to AD	Kan	This study
pSUPROM MTB 243ext-Rieske-Bla	pSUPROM	pSUPROM MTB Rieske-Bla extension of sequence from aa 227 to 243	Kan	This study

2.4.3 Amplification of DNA by Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction (PCR) is a biochemical technique that allows the amplification of specific regions of DNA by several orders of magnitude. This procedure utilises thermostable DNA polymerase enzymes and oligonucleotide primers, which are designed to be complementary to the 5' and 3' ends of a thermally denatured single stranded DNA template. DNA elongation occurs in a 5' to 3' direction, enabling the strands to extend towards one another.

A typical PCR reaction mixture was composed of optional DMSO, less than 50 ng template DNA (or as required), 1 µl of forward/reverse primer (from a 100 µM stock), 0.5 µl dNTPs (from a 20 mM stock), 0.5 µl DNA Polymerase and appropriate enzyme buffer. This was made up to 50 µl using ultra-pure water. The PCR programme consists of three essential steps; a denaturing step at 96°C, an annealing step at 50-60°C and an elongation step at 72°C. These steps were repeated for 15-30 cycles and followed by a final elongation at 72°C. PCR products were assessed by agarose gel electrophoresis and were cleaned using a QIAquick® Gel extraction kit or a QIAquick® PCR Purification kit (Qiagen) before use in subsequent cloning applications.

2.4.4 Site-directed mutagenesis by Quikchange™ PCR

Site-specific mutations were introduced based on the Quikchange™ manual from Stratagene. Complementary primers were designed (Sigma) to introduce specific amino acid substitutions into a template sequence. The codon of interest was positioned in the middle of the 30 base pair sequence. Truncation and deletion mutations were generated using Liu and Naismith's modified Quikchange™ method of partial overlapping primer design (Liu & Naismith, 2008). DNA was amplified by PCR before the methylated template DNA present in the sample was digested with *Dpn*I (Roche). DNA was subsequently transformed into chemically competent cells of *E. coli* strain DH5α and extracted by use of QIAPrep® Spin Miniprep Kit (Qiagen) according to the

manufacturer's instructions (Section 2.4.2). All oligonucleotides used in this study are found in the Appendix.

2.4.5 Agarose gel electrophoresis

DNA samples were analysed by agarose gel electrophoresis using 1% (w/v) agarose gels prepared with TAE buffer and containing 0.001% (v/v) Gel Red dye (Biotium). Loading dye was added to samples to create a visible running front and DNA size markers (1 kb Plus DNA Ladder, Invitrogen) were run on gels to allow identification of target bands.

2.4.6 DNA digestion and preparation for cloning

DNA was digested using 10 units of the appropriate restriction enzymes with the coordinated restriction enzyme buffer. This was in a final volume of 40 or 50 µl depending on whether the digestion was 1 µg of plasmid or 50 µl of PCR product. Samples were incubated for at least 2 h at 37°C. Digested vectors were then treated with alkaline phosphatase to dephosphorylate the vector thereby preventing self-religation. This involved incubation with 5 µl of 10x alkaline phosphatase buffer and 5 µl alkaline phosphatase at 37°C for 30 min. Samples were then assessed using agarose gel electrophoresis and visualised under UV light. Bands of interest were excised and purified using a QIAquick® Gel extraction kit (Qiagen) according to the manufacturer's instructions.

2.4.7 DNA ligation

DNA fragments of interest were ligated to a linearized vector using T4 DNA ligase (Roche). A molar excess ratio of vector:insert of 1:3 was used. The following equation (1) was used to calculate the appropriate amounts of DNA.

$$mass_{vector}[ng] = \frac{mass_{vector}[ng] \times size_{insert}[bp]}{size_{vector}[bp]} x3 \quad (1)$$

DNA fragments were ligated in a final volume of 10 µl containing ligation buffer and 1 µl of ligase. Ligations were performed at 18°C for 1 h or overnight before the entire reaction volume was used in transformation with chemically competent cells of *E. coli* strain DH5α.

2.4.8 DNA sequencing

Subsequent to cloning, DNA sequencing was used to confirm sequences of plasmids. Sequencing of DNA was performed by the MRC PPU DNA Sequencing & Services at the School of Life Sciences, University of Dundee. Chromatogram files obtained from sequencing reactions were analysed using BioEdit Sequence Alignment Editor (Copyright © 1997-2011 Tom Hall).

2.5 Protein methods

2.5.1 SDS-PAGE

SDS polyacrylamide gel electrophoresis (SDS-PAGE) is used to separate proteins under denaturing conditions according to molecular weight (Laemmli, 1970). Tris-glycine SDS-PAGE gels of 0.75mm thickness were prepared for use with the Mini-PROTEAN II system (Bio-Rad) at varying polyacrylamide concentrations (7.5%, 12% or 15%) depending on the experimental requirements. The compositions of resolving and stacking gels can be found in Table 2.7.

Table 2.7 Composition of SDS-PAGE resolving and stacking gel.

Resolving gel		**Stacking gel**	
Acrylamide/bis-acrylamide (37.5:1)	7.5% or 12% or 15% (v/v)	**Acrylamide/bis-acrylamide (37.5:1)**	4% (v/v)
Tris-HCl pH 8.8	0.375 M	**Tris-HCl pH 6.8**	0.125 M
SDS	0.1% (w/v)	**SDS**	0.1% (w/v)
APS	0.1% (w/v)	**APS**	0.1% (w/v)
TEMED	0.1% (v/v)	**TEMED**	0.1% (v/v)

Tris-glycine gels were submerged in a gel electrophoresis tank (Bio-Rad) filled with SDS running buffer. Samples were prepared by mixing 1:5 with 6x sample buffer or 1:1 with 2x Laemmli sample buffer both containing β-mercaptoethanol and loaded along with the marker, Precision Plus Protein™ All Blue standards (Bio-Rad). The gel was run at 100 V

until the dye passed beyond the stacking gel, before the voltage was increased to 150-200 V until the appropriate marker reached the bottom of the gel. Gels were then used further in semi-dry or dry western immunoblotting.

2.5.2 Semi-dry and dry Western Blotting

Samples to be analysed were first separated using SDS-PAGE as described above, subsequently they were transferred to nitrocellulose membrane using either the semi-dry or dry transfer systems.

For semi-dry transfer, gels were soaked in Tris-glycine transfer buffer before being placed on pre-soaked nitrocellulose membrane (Amersham Hybond-ECL, GE Healthcare). These were sandwiched between 4 pieces of pre-soaked 3MM Whatmann paper before being placed inside a TransBlot SD SemiDry Transfer Cell (Bio-Rad) for protein transfer. Transfer was performed at 10 V for 30 min.

The iBlot2 device (Life technologies) was used for dry transfer along with pre-manufactured iBlot transfer stacks, a top cathode stack and a bottom anode stack, the transfer was performed according to the manufacturer's instructions. The Top cathode stack is made of a gel matrix containing cathode buffer and a copper cathode layer and the bottom anode stack is made of a copper anode layer, a gel matrix containing anode buffer and a nitrocellulose blotting membrane. The bottom stack was placed on the machine topped with the pre-run SDS-PAGE gel, a deionised water soaked filter paper, the top stack and a sponge. Transfer programme p0 (20 V for 1 min, 23 V for 4 min and 35 V for 2 min) was used to transfer proteins onto nitrocellulose membrane.

After transfer the membrane was blocked in TBS-Tween with 5% skimmed milk for 1 h at room temperature or overnight at 4°C, with shaking. This was followed by washing in TBS-Tween, before a 1 h incubation with primary antibody suitably diluted in TBS-Tween. This was followed by 3 washes in TBS-Tween before the incubation for 1 h with the secondary antibody, also suitably diluted in TBS-Tween. The working solutions for antibodies are provided in Table 2.8. Enhanced chemiluminescent detection was

used to visualise the bands. The secondary antibodies (and some primary antibodies) were conjugated to horseradish peroxidase (HRP) this oxidises luminol in the presence of peroxide, resulting in the emission of light proportional to protein quantities as the luminol decays. This reaction allows for the detection of HRP-coupled antibodies bound directly or indirectly to the target protein on the membrane. These immunoreactive protein bands were detected using Clarity™ Western ECL substrate kit (Bio-Rad) and exposure to medical film (Konica Minolta) developed in the medical film processor SRX-101A (Konica Minolta) or visualised with a CCCD camera (GeneGNOME XRQ Syngene).

Table 2.8 Antibodies used in this study.

Primary Antibody	Working concentration	Raised in	Reference
Monoclonal Rieske peptide antibody	1:10000	Rabbit	Generated by GenScript *
Monoclonal Anti-HA peroxidase	1:3000	Mouse	Sigma-aldrich (Cat. #H6533)
Polyclonal Rieske antiserum	1:10000	Rabbit	(Keller *et al.*, 2012)
Monoclonal Bla antibody	1:5000	Mouse	Abcam® (Cat. #ab12251)
Monoclonal BamA antibody	1:10000	Rabbit	(Lehr *et al.*, 2010)
Secondary Antibody			
anti-mouse IgG HRP conjugate	1:10000	Goat	Bio-rad
anti-rabbit IgG HRP conjugate	1:10000	Goat	Bio-rad

* N-terminal epitope (sequence: CLPPHEPRVQDVDER)

2.5.3 Determination of protein concentration

Protein concentrations in protein samples were determined by the Lowry method (Lowry *et al.*, 1951) using the DC™ Protein Assay kit (Bio-Rad). A standard curve was generated using Bovine serum albumin (BSA) for sample comparison.

2.5.4 Protein precipitation

Acetone precipitation was used in this work as it is a good method for precipitating soluble proteins by removing salts and lipid-soluble contaminants. Ice cold acetone was added to samples at a volume four times the amount of sample, this mixture was then vortexed and incubated for 1 h at -20°C. The samples were then centrifuged for 10 min

at 16000 x *g* and the supernatant was discarded. The protein pellet was air dried for 20 min at room temperature to remove the remaining acetone before the pellet was resuspended in the appropriate buffer.

2.6 Preparation of membrane fractions

50 µl of a stationary phase culture was subcultured into 50 ml LB and grown at 37°C with shaking until OD_{600} of 0.2 was reached. The cells were harvested by centrifugation at 3000 x *g* for 10 min at 4°C and the supernatant was discarded. The cell pellet was resuspended in 1 ml of Buffer 1. The cells were lysed by sonication at 20% amplitude for 1 min (Branson Digital Sonifier). The lysates were then centrifuged to remove cell debris at 17000 x *g* for 5 min followed by ultracentrifugation to pellet the membrane at 227000 x *g* for 30 min at 4°C. The pellets were resuspended in Buffer 2. An appropriate amount of 6x sample buffer with 5% β-mercaptoethanol was added before the samples were analysed by SDS-PAGE and Western blotting.

2.7 Densitometry analysis of protein production

Image Processing and Analysis in Java (ImageJ, NIH, (Schneider *et al.*, 2012)) was used to analyse the protein level of Rieske fused variants. Membrane fractions were prepared as Section 2.6 and analysed by SDS-PAGE and Western Blotting with anti-Rieske-peptide to detect the Rieske protein, and anti-BamA as a loading control. The density of the band was calculated by densitometry analysis and Rieske-associated signals were normalised with BamA-associated signals. The results were expressed as percentage of the normalised signal obtained for 'wild type' Rieske-AmiA or Rieske-Bla.

2.8 Sulphydryl labelling of cysteine residues

Labelling of cysteine residues was performed with methoxypolyethylene glycol maleimide (mal-PEG). Mal-PEG has a molecular mass of 5 kDa and forms a covalent 3-thiosuccinimidyl ether linkage with thiols *via* the maleimide group, this reaction is

specific for pH 6.5-7.5. This coupling of cysteine residues with mal-PEG can be detected by a band shift after SDS-PAGE and Western blotting. Mal-PEG is membrane-impermeable, therefore, after treatment only accessible cysteines are labelled. Single cysteine substituted cysteine variants of Rieske were expressed in appropriate *E. coli* strains and labelling was undertaken. Sulphydryl labelling was performed on crude membrane fractions where all cysteines should be accessible and whole cells where only those cysteines in the periplasm should be labelled.

2.8.1 Sulphydryl labelling in membrane fractions

Membrane fractions were prepared as described in Section 2.6. Sulphydryl labelling of membrane fractions was carried out using membrane fractions resuspended in HEPES/$MgCl_2$ buffer (pH 6.8). 30 µl of membrane fraction was incubated with 5 mM mal-PEG in a final volume of 50 µl for 1 h at room temperature, the reaction composition is described in Table 2.9.

Table 2.9 Composition of sulphydryl labelling reaction with crude membrane fractions.

	Unlabelled	labelled	[C]f
HEPES buffer	20 µl	15 µl	
Membrane fraction	30 µl	30 µl	
mal-PEG 50mM		5 µl	5 mM
Final volume	50 µl	50 µl	

The labelling reaction was stopped by the addition of 0.5 M dithiothreitol (DTT, 100 mM final concentration). The samples were mixed with 50 µl 2x Laemmli sample buffer containing 5% β-mercaptoethanol and analysed by SDS-PAGE and Western blotting.

2.8.2 Sulphydryl labelling in intact cells and membrane preparation

1500 µl of a stationary phase culture was subcultured into 50 ml LB with appropriate antibiotics and grown at 37°C with shaking until OD_{600} of 0.6-0.7 was reached. The cells were harvested by centrifugation at 3000 x *g* for 15 min and the supernatant was

discarded. The cell pellet was resuspended and washed in 20 ml HEPES/$MgCl_2$ buffer. The cells were spun down at 3000 x *g* for 15 min and the supernatant was discarded. The cell pellet was resuspended in 1ml HEPES/$MgCl_2$ buffer (pH 6.8). per 0.3 units of OD_{600} measured before harvesting. The resuspended cells were split into two 800 µl aliquots supplemented with either buffer or mal-PEG in a final volume of 1 ml. EDTA was included in all of the samples to increase the permeability of mal-PEG through the OM (Leive, 1968). The reaction composition is described in Table 2.10.

Table 2.10 Composition of sulphydryl labelling reactions with whole cells.

	unlabelled	labelled	[C]f
HEPES buffer	180 µl	80 µl	
EDTA 250mM	20 µl	20 µl	5mM
Cells	800 µl	800 µl	
PEG-Mal 50mM		100 µl	5mM
Final volume	1 ml	1 ml	

After incubation for 1 h at room temperature the reactions were stopped with 1M DTT (100 mM final concentration). 100 µl of each sample was precipitated as described in Section 2.5.4 and resuspended in 30 µl of Buffer 2. The remaining 900 µl of each sample was used for membrane preparation. 300 µl of HEPES/$MgCl_2$ buffer was added to each sample to decrease the viscosity. The samples were then lysed by sonication at 20% amplitude in 5 s bursts for a total of 1 min (Branson Digital Sonifier). The lysates were centrifuged to remove cell debris at 17000 x *g* for 5 min followed by ultracentrifugation to pellet the membranes at 227000 x *g* for 30 min at 4°C. The resulting membrane pellets were then resuspended in Buffer 2. An appropriate volume of 6x sample buffer with 5% β-mercaptoehanol was added before the samples were analysed by SDS-PAGE and Western blotting.

2.9 Growth assays

Two fusion reporters were used throughout this study; AmiA and Bla. Growth assays were developed for each fusion reporter to test for Tat- or Sec-dependent transport using a *tat*$^{+}$ or *tat*$^{-}$ strain accordingly.

2.9.1 Growth assays for fusion reporter AmiA

Growth assays to test Tat-dependent survival in the presence of the detergent SDS were performed in two ways. Stationary phase cultures were subcultured and grown at 37°C until an OD_{600} of 0.5 was reached, after which 5 µl of culture was spotted on LB solid medium supplemented with 2% SDS and appropriate antibiotics. Alternatively 5 µl aliquots of a stationary phase culture was spotted on LB solid medium supplemented with 1% SDS as a serial dilution of 10^4 cells to 10 cells per 5 µl. Plates were incubated at 37°C for 16 h.

2.9.2 Growth assays for fusion reporter Bla

Growth assays to test Sec/Tat-dependent survival in the presence of ampicillin were performed in two ways. A 5 µl aliquot of a stationary phase culture was spotted on LB solid medium supplemented with increasing concentrations of ampicillin from 0-50 µg/ml, these were incubated at 37°C for 16 h. Alternatively a growth curve was performed where stationary phase cultures were diluted to OD_{600} of 0.03 in LB with increasing concentrations of ampicillin from 0-50 µg/ml, then 100 µl culture was added to a 96 well plate. This was placed into Synergy 2 platereader (Biotek) and incubated at 37°C with continuous shaking. OD_{600} readings were taken every 20 min for 10 h.

Bla fusion variants were tested for Sec-dependent transport using the *tat*$^{-}$ strain (DADE) and a quantitative assay that determines the antimicrobial susceptibility of the bacteria to ampicillin *via* the Minimum Inhibitory Concentration (M.I.C.). To do this Oxoid M.I.C.Evaluator™ (Thermo Fisher Scientific) test strips were used, these contain a continuous antibiotic concentration gradient, in this case ampicillin with a gradient from 0-256 µg/ml. The strip is placed on an LB agar plate containing a lawn of bacteria and

after incubation, a symmetrical inhibition ellipse centred along the strip is seen. The M.I.C. value (in µg/ml) is read from the scale where the pointed end of the ellipse intersects the strip.

Stationary phase cultures were diluted to OD_{600} of 0.1 and an LB agar plate was inoculated by swabbing with the diluted culture to generate a lawn of bacteria. The M.I.C.Evaluator™ test strip was placed onto the lawn of bacteria and the plates were incubated at 37°C for 18 h after which the M.I.C. was read off the strip. A minimum of 3 biological replicates per variant was tested.

3 Development of fusion reporter assays

3.1 Introduction

3.1.1 The use of fusion reporters in protein analysis

Keller *et al.* (2012) used reporter fusions to determine the topology of the transmembrane region of the *Streptomyces coelicolor* Rieske FeS protein when it was expressed in *E. coli*. Gene fusions have been used extensively as tools to probe transport pathways and determine transmembrane protein topologies. The basis of this approach is to choose a reporter protein with a specific subcellular compartmental activity which is then fused to protein fragments containing prospective targeting sequences. A good fusion reporter requires a distinct well-defined phenotype, these generally involve plate screens which specify a growth output or a colour change. Consequently, when this fusion undergoes testing, the targeted compartment is indicated by the activity of the reporter protein (Stanley *et al.*, 2002). The following proteins will be discussed further in this section.

3.1.1.1 Amidase A

E. coli contains three N-acetylmuramyl-$_L$-alanine amidases; AmiA, B and C. These are periplasmic enzymes that cleave the peptide moiety from N-acetylmuramic acid thereby removing murein crosslinks. This is particularly important during cell division as splitting the murein septum is vital to separation of the daughter cells (Bernhardt & de Boer, 2003, Heidrich *et al.*, 2001, Uehara *et al.*, 2010). Deletion of these amidases results in cells forming long chains, and similar cell chains along with a pleiotropic OM defect are also observed when essential Tat system components are absent (Stanley *et al.*, 2001). It was subsequently shown that AmiA and AmiC are exported to the periplasm by the Tat pathway and the failure to export these proteins accounts for the cell-chaining and OM defect of *tat* mutant strains (Heidrich *et al.*, 2001, Ize *et al.*, 2003).

Due to the cell envelope defects, deleting these Tat-dependent enzymes in *E. coli* strains causes the cells to be sensitive to killing by the detergent SDS (Ize *et al.*, 2003). This

makes for a good phenotypic screening system, as strains with an inactive Tat system or *tat*$^{+}$ strains lacking AmiA and AmiC in the periplasm will fail to grow on SDS-containing growth media (Keller *et al.*, 2012).

3.1.1.2 β-Lactamase

β-lactam antibiotics, like ampicillin, kill cells by inactivating the enzymes involved in cell wall biosynthesis, and β-lactamases (Bla) cleave the amide bond in the β-lactam ring of the inbound antibiotics, thereby protecting cells against lysis (Broome-Smith & Spratt, 1986). Bla is required to be in the periplasm for function and although it is usually a substrate of the Sec pathway, it is also compatible with export by the Tat pathway if it is supplied with a twin-arginine signal peptide (Pradel *et al.*, 2009, McCann *et al.*, 2007). As this protein specifies ampicillin resistance, it has frequently been used as a reporter, with phenotypic screening of growth on ampicillin-containing medium when Bla has been translocated into the periplasm (Stanley *et al.*, 2002, McCann *et al.*, 2007, Broome-Smith *et al.*, 1990, Mansell *et al.*, 2010).

3.1.1.3 Maltose binding protein

E. coli maltose binding protein (MalE or MBP) is an essential periplasmic component of a membrane bound transporter ATPase binding cassette superfamily which catalyses the uptake of maltose. Its export to the periplasm is *via* the Sec pathway, with a strong dependence on the chaperone, SecB (Fekkes & Driessen, 1999). However, replacement of its signal peptide with a Tat-targeting sequence can specifically redirect MBP to the Tat pathway, although some residual Sec translocation of MBP may still be observed depending on the Tat targeting sequence used (Blaudeck *et al.*, 2003, DeLisa *et al.*, 2002). *malE*$^{-}$ cells expressing a plasmid that produces a low proportion of MBP, which is exported to the periplasm, can grow on minimal media if provided with a maltose carbon source, this provides a straightforward qualitative method for detecting MBP export.

The leaky cell envelope phenotype of *tat*$^-$ strains also effects the utility of MBP as an export reporter, as even in the absence of *malE,* some maltose can be transported by *tat* mutants (but not by *malE*$^-$ *tat*$^+$ strains). This impacts on the utility of this reporter when testing for growth on maltose-containing minimal medium as false positives may be found. High concentrations of calcium or magnesium have been found to act as a chemical repair to this pleiotropic OM defect of *tat*$^-$ cells (Caldelari *et al.*, 2008).

3.1.2 Co-ordinating crosstalk between the Sec and Tat machineries during assembly of Rieske

An alignment of amino acid sequences of Rieske FeS proteins from different bacteria shows the highly conserved regions between the different Rieske proteins (Figure 1.19). This includes the essential twin-arginine motif required for translocation by the Tat machinery, which is found N-terminal to TMD3 in actinobacterial Rieske proteins (*S. coelicolor, C. glutamicum* and *M. tuberculosis)* (Figure 1.19, highlighted in purple). This highly conserved motif strongly suggested that despite the unusual topology of these Rieske proteins they still use the Tat system for translocating the globular FeS domain across the membrane.

Keller *et al.* (2012) investigated the Tat-dependence of the membrane portion of *S. coelicolor* Rieske using fusion reporter proteins where the TMDs of Rieske were fused to MBP or AmiA in order to assess the translocation of TMD3 of Rieske (as indicated by the arrow in Figure 1.19). The reporter assays used were able to identify translocation of the fused reporter into the periplasm leading to the conclusion that *S. coelicolor* Rieske was inserted into the membrane using a co-operation of the Sec and Tat translocation pathways.

The overall aim of this thesis was to define key regions of *S. coelicolor* Rieske that are important for the interaction of the protein with the Sec and Tat machineries. To this end it is essential that the TMDs of Rieske are fused to reporter proteins that provide robust

and reliable assays, and that can indicate clear positive and negative results. Therefore, the aim of the work in this chapter was to test and validate reporters to assess Sec and Tat translocation of Rieske and to develop quantitative methods to determine the degree of Sec and Tat integration. This would ultimately support a full programme of mutagenesis aimed at dissecting features of Rieske that dictate interaction with the Sec and Tat pathways.

3.2 Results

3.2.1 Selection of reporter fusions to reliably test Sec- and Tat-dependence of Rieske variants

In previous work (Keller *et al.*, 2012) two reporter proteins were fused after TMD3 of *S. coelicolor* Rieske; MBP and AmiA. These constructs were expressed in *E. coli tat*$^+$ and *tat*$^-$ (Δ*tatABC*/Δ*tatE*) strains and different assays were undertaken to test for Tat dependence. The MBP assay relies on the translocation of MBP to the periplasmic side of the membrane where it promotes transport and fermentation of the maltose causing an acidification reaction, which is detected in liquid medium by a pH indicator dye and a colour change from purple to yellow. In comparison, the AmiA assay relies on the translocation of AmiA into the periplasm allowing the growth of cells lacking native AmiA and AmiC on solid media containing SDS.

The *E. coli tat*$^-$ mutant strain is known to have a pleiotropic cell envelope phenotype which leads to a leaky OM (Ize *et al.*, 2003, Stanley *et al.*, 2001). This phenotype can be partly compensated by the addition of $CaCl_2$ but the envelope phenotype cannot be completely repaired and confers a partial *mal*$^+$ phenotype even in the absence of periplasmic MBP (Caldelari *et al.*, 2008). The combination of the leaky membrane phenotype of the *E.coli tat*$^-$ strain and the indirect readout of a colour change makes it very difficult to score for *mal*$^+$/*mal*$^-$ phenotypes in the *tat*$^-$ strain. To undertake a mechanistic dissection of *S. coelicolor* Rieske membrane insertion, a robust reporter system is required where the results are to be reliable and small changes can be scored, ruling out the utility of MBP as a reporter fusion protein.

Unlike MBP, AmiA is a more reliable fusion reporter in this context. When expressed in a *tat*$^+$ strain lacking chromosomally-encoded AmiA/C, AmiA will be translocated to the periplasm allowing cells to grow on media containing SDS. If AmiA is not translocated, for example when expressed in the *tat*$^-$ strain, there will be no growth on media containing

SDS. As AmiA has a direct and reliable readout it was selected as one of the fusion reporters in this project.

A different reporter that has previously been used in fusion reporter studies is Bla (Lee & Hughes, 2006, Stanley *et al.*, 2002). As Bla specifies ampicillin resistance, translocation into the periplasm will allow growth of cells on medium containing ampicillin, therefore, a direct and reliable readout can be assessed. This would potentially make Bla a robust fusion reporter for evaluating Rieske insertion.

Keller *et al*. (2012) used the pBAD24 vector to generate MBP fusion constructs, which expresses genes that are under the control of the arabinose promoter (Guzman *et al.*, 1995). This plasmid specifies ampicillin resistance and thus cannot be used in conjunction with the Bla fusion reporter which also specifies ampicillin resistance. Therefore a different vector system, pSUPROM, was chosen for cloning of DNA encoding the fusion reporter proteins, which specifies kanamycin resistance and expresses genes under the control of the *tat* promoter (Jack *et al.*, 2004). To maintain consistency between reporter fusion proteins the same expression vector was used to express both AmiA and Bla fusions with the *S. coelicolor* Rieske protein.

The generation of the fusion reporters is described in detail in Chapter 2 and followed the approach of Keller *et al*. (2012), whereby the globular FeS domain was genetically removed from *S. coelicolor* Rieske and either the mature region of the *E. coli* Tat substrate AmiA (to generate Rieske-AmiA) or the mature region of Bla obtained by PCR amplification from pBR322 (to generate Rieske-Bla) were fused to the C-terminal end of the TMD3 of Rieske, as outlined in Figure 3.1. As in Keller *et al*. (2012), the signal peptidase I cleavage site of AmiA (Ala-Xaa-Ala) was included in the construct, thus allowing release of the amidase from the membrane.

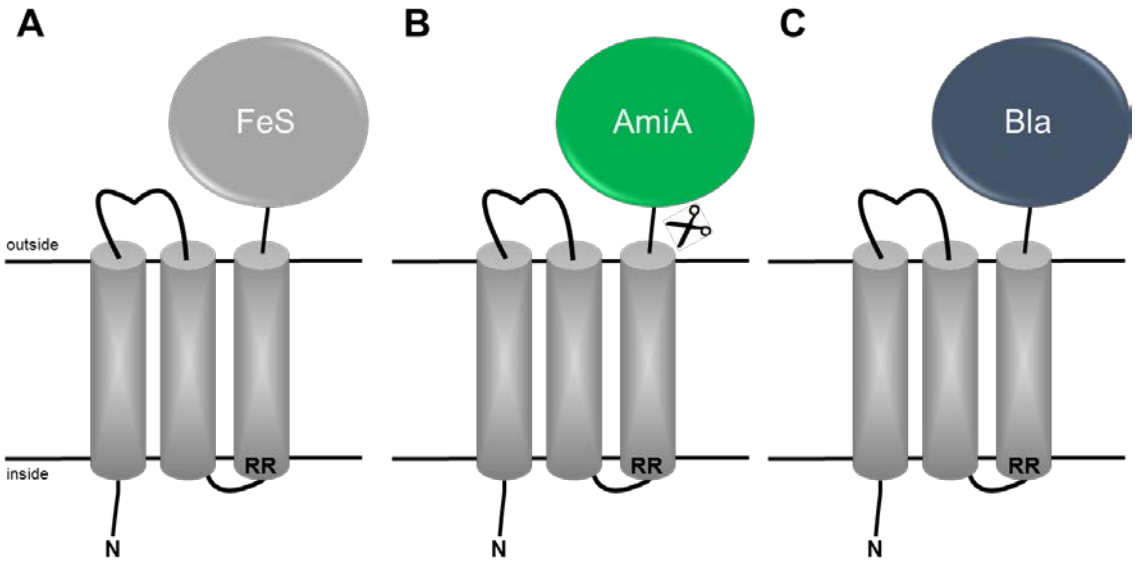


Figure 3.1 Generation of *S. coelicolor* Rieske TMD fusion proteins with reporters AmiA and Bla. (A) Schematic representation of *S. coelicolor* Rieske comprising of three TMD followed by the globular FeS (FeS) cofactor-containing domain. (B) Fusion protein Rieske-AmiA where the FeS domain has been replaced with mature *E. coli* AmiA domain (i.e. with the native N-terminal twin-arginine signal sequence removed). A cleavage site has also been included between the C-terminal end of TMD3 and the AmiA sequence to allow release AmiA of to the periplasm to better access the peptidoglycan substrate (indicated by scissors). (C) Fusion protein Rieske-Bla where the FeS domain has been replaced by mature β-lactamase (Bla) domain with the native N-terminal Sec signal sequence removed. The position of the twin-arginine motif is indicated by RR at the N-terminal end of TMD3.

3.2.2 Assessing Tat-dependence of the Rieske-AmiA Reporter Fusion

3.2.2.1 SDS spot testing shows Tat-dependent membrane insertion of Rieske-AmiA

The fusion construct Rieske-AmiA was tested for Tat-dependence using *E. coli tat*$^+$ and *tat*$^-$ strains, MCDSSAC (Ize *et al.*, 2003) and MCDSSACΔtat (Keller *et al.*, 2012), respectively. These strains have chromosomal deletions in the AmiA and AmiC signal peptide coding regions (Δ*ssAmiAC*), thereby trapping the expressed proteins in the cytoplasm where they are unable to function, and the *tat*$^-$ strain lacks the *tatABC* genes. The *tat*$^+$ and *tat*$^-$ strains harbouring plasmid-encoded Rieske-AmiA, a positive control of plasmid-encoded full length AmiA (with its native Tat signal sequence) and a negative control of pSUPROM (Empty vector) were grown in liquid culture lacking SDS and were subsequently spotted onto LB medium containing 2% SDS. As Figure 3.2A indicates, translocation of Rieske-AmiA into the periplasm is Tat-dependent, as growth occurred in

the presence of SDS when the construct was expressed in the *tat*$^+$ strain, but there was no growth when the Rieske-AmiA construct was expressed in the *tat*$^-$ strain. As expected, the positive control, full length AmiA, showed typical Tat-dependent translocation behaviour. This result confirms the previous work of Keller *et al*. (2012) showing that AmiA is a useful reporter to study Tat-dependence of Rieske membrane insertion (Figure 3.2B).

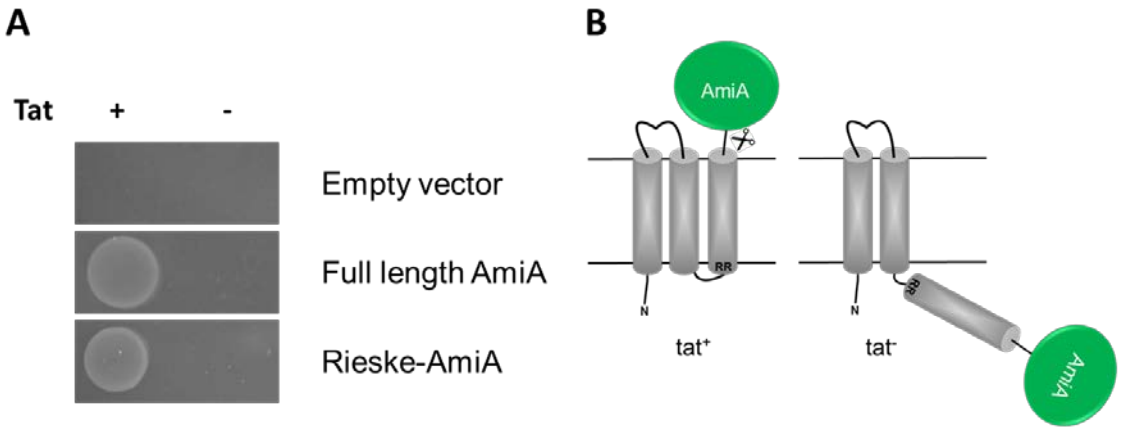


Figure 3.2 Tat-dependent translocation of the Rieske-AmiA fusion protein.
(A) Spot tests of strains MCDSSAC (*tat*$^+$ strain) and MCDSSACΔtat (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), or the pSUPROM vector encoding either full length AmiA (with its native Tat signal peptide) or Rieske-AmiA. Strains were grown overnight in LB medium, subcultured to give a final OD_{600} of 0.5 and 5 µl aliquots were spotted onto LB medium containing 2% (w/v) SDS. LB agar plates were incubated at 37°C for 16h. (B) Predicted topology of the Rieske-AmiA fusion protein when expressed in *tat*$^+$ and *tat*$^-$ strains.

3.2.2.2 A Rieske-AmiA twin-lysine variant is still recognised by the Tat machinery

The next step was to generate a substitution in the twin-arginine motif of the Rieske-AmiA variant to abolish Tat-translocation. The twin-arginine motif in Rieske-AmiA is located at the N-terminal end of TMD3 (Figure 1.19), and substitutions of the consensus twin-arginines are the most effective way to prevent translocation of proteins through the Tat machinery. In most cases translocation through the Tat pathway can be fully abolished by a double arginine to lysine substitution (Bachmann *et al.*, 2006, Buchanan *et al.*, 2001, Widdick *et al.*, 2008, Stanley *et al.*, 2000), and indeed this was shown previously to block recognition of Rieske-AmiA by the Tat machinery (Keller *et al.*, 2012). The double mutation of twin-arginine to twin-lysine (RRKK) in the Tat motif of

Rieske-AmiA was generated in the pSUPROM construct and introduced into the *tat*$^+$ and *tat*$^-$ strains before undertaking the spot assay on media containing SDS. Surprisingly, and in contrast to Keller *et al.* (2012), the *tat*$^+$ strain producing the RRKK substituted Rieske-AmiA was able to grow in the presence of SDS (Figure 3.5).

In most soluble native Tat substrate proteins the RRKK substitution almost universally blocks Tat translocation, even if the substrate is highly overexpressed (e.g. (Sargent *et al.*, 2001)). However, mechanistically, the twin-lysine substitution must have some compatibility with the Tat pathway because export of very sensitive Tat reporters such as colicin V and MBP can still be detected when twin-lysines are introduced into the signal peptide (Kreutzenbeck *et al.*, 2007, Ize *et al.*, 2002a, Ize *et al.*, 2002b). AmiA is, however, not as sensitive a reporter as colicin V and MBP, and indeed a twin-lysine substitution of the twin-arginines of the native AmiA signal peptide completely blocked detectable transport by the Tat pathway, even when the protein was overproduced (Ize *et al.*, 2003). It should be noted, however, that the Rieske-AmiA fusion protein is already anchored at the membrane due to insertion of the first two TMD by the Sec pathway. This effectively constrains diffusion of the protein to two dimensions instead of three, greatly increasing the concentration of this protein close to the Tat machinery, potentially facilitating recognition of the variant twin-lysine signal sequence.

3.2.2.3 The Rieske-AmiA fusion protein is produced at a higher level from pSUPROM than from pSU18

As described above, the variant RRKK-Rieske-AmiA was able to support growth of the MCDSSAC strain on media containing SDS. This is in contradiction to the work by Keller *et al.* (2012) where export of RRKK-Rieske-AmiA was not observed in the same strain background. To examine this further, the growth tests were repeated using the RRKK-Rieske-AmiA construct made by Keller *et al*. (2012) in the pSU18 vector and the RRKK-Rieske-AmiA generated here in the pSUPROM vector. As shown in Figure 3.3, no growth on SDS was observed when the variant fusion protein was produced from

pSU18, in agreement with the published findings, whereas clear growth was again observed when it was produced from pSUPROM.

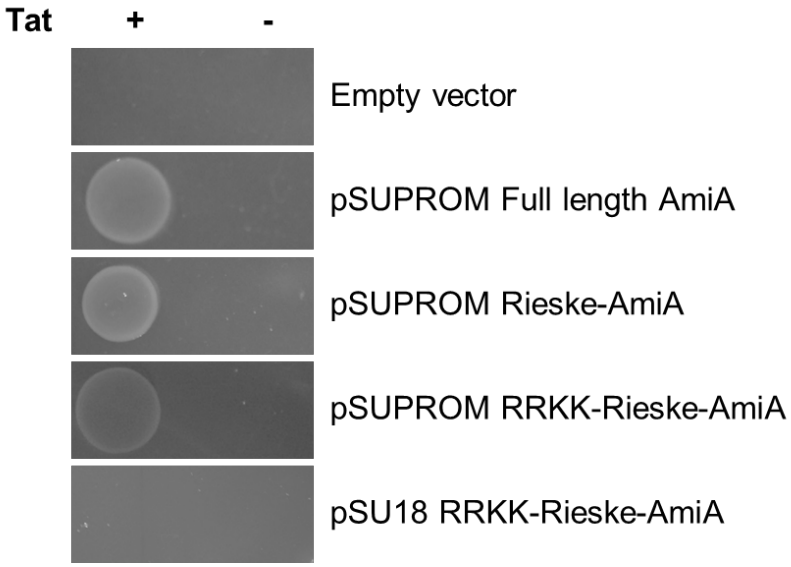


Figure 3.3 Growth in the presence of SDS depends upon the vector from RRKK-Rieske-AmiA is produced.
Spot tests of strains MCDSSAC (*tat*$^+$ strain) and MCDSSACΔtat (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), or the pSUPROM vector encoding full length AmiA, 'wild type' Rieske-AmiA and twin-lysine substituted Rieske-AmiA (RRKK-Rieske-AmiA), or the pSU18 vector encoding the twin-lysine substituted Rieske-AmiA (RRKK-Rieske-AmiA). Strains were grown overnight in liquid media, subcultured to give a final OD_{600} of 0.5 and 5 µl aliquots were spotted onto LB medium containing 2% (w/v) SDS. LB agar plates were incubated at 37°C for 16 h.

The major difference between the two constructs is the nature of the plasmid vector, with Keller *et al*. (2012) using plasmid pSU18 as an expression vector whereas the work in this thesis uses the pSUPROM expression vector. pSUPROM is a standard expression vector, based on the kanamycin resistance plasmid pSU40 (Bartolome et al., 1991). This plasmid, which contains the *lac* promoter, has been modified to also contain the *E. coli* constitutive *tatA* promoter followed by a strong ribosome binding site (RBS) – this enables a relatively high level of gene expression (Jack *et al.*, 2004). In comparison pSU18 contains the *lac* promoter followed by a weak RBS (Bartolome *et al.*, 1991). It should be noted that all of the strains used in this thesis and in the study of Keller *et al*. (2012) are MC4100 derivatives which are *lacI*$^-$, and therefore the *lac* promoter is constitutively active in these strain backgrounds.

To determine if the growth difference seen in Figure 3.3 results from in a difference in the level of protein production between the two vectors, cells from the *tat*$^+$ strain harbouring both pSU18 RRKK-Rieske-AmiA and pSUPROM RRKK-Rieske-AmiA were analysed by Western blotting with an anti-Rieske antiserum. As shown in Figure 3.4, the level of Rieske-AmiA fusion protein produced from pSU18 RRKK-Rieske-AmiA is much lower than from pSUPROM RRKK-Rieske-AmiA. This may be a reflection of the double promoter present on pSUPROM and the fact that pSU18 has a weaker RBS. Thus it can be concluded that the ability of the pSUPROM expressed variant of RRKK-Rieske-AmiA to support growth of the MCDSSAC strain on media containing SDS is a result of the elevated level of protein production relative to that produced from pSU18 RRKK-Rieske-AmiA.

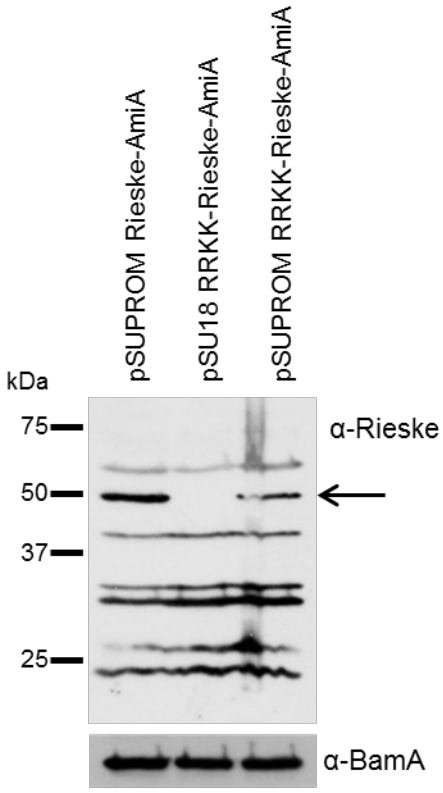


Figure 3.4 RRKK-Rieske-AmiA is produced at different levels from pSU18 and pSUPROM.
Equivalent amounts of whole cells of the *tat*$^+$ strain MCDSSAC harbouring pSUPROM Rieske-AmiA, pSUPROM RRKK-Rieske-AmiA or pSU18 RRKK-Rieske-AmiA were separated by SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Rieske (top) and anti-BamA (bottom) antisera. BamA was used as a loading control and bands corresponding to Rieske-AmiA are indicated by an arrow.

3.2.2.4 Identifying inactive twin-arginine variants of Rieske-AmiA

Since a variant of Rieske-AmiA that was unable to engage with the Tat pathway was required, further substitutions of the twin-arginine motif were generated to find a mutation that completely abolished Tat translocation of Rieske-AmiA. Therefore, substitutions of the twin-arginines to lysine-glutamine (RRKD), twin-alanine (RRAA), alanine-aspartate (RRAD) and deletion of both arginines (ΔRR) were generated and the variants were expressed in the *tat*$^{+}$ and *tat*$^{-}$ strains and tested for growth on media containing SDS.

Figure 3.5 shows that mutation of the twin-arginine to lysine-glutamine still retained some recognition by the Tat pathway. This contrasts with the findings of Kreutzenbeck *et al.* (2007) who showed that a similar substitution to the twin-arginines of a TorA signal sequence-MalE reporter fusion abolished Tat-translocation. However, mutations of the twin-arginine motif to di-alanine, alanine-glutamine or deletion of both arginines successfully abolished Tat translocation. It is believed that the Tat machinery to some extent relies on electrostatic interactions to recognise the twin-arginine in the Tat motif. This is supported by the results presented in Figure 3.5 as the abolition of positively-charged residues at the twin-arginine positions of Rieske-AmiA (by substitution for the electro-neutral twin alanine, or negatively charged alanine-aspartate pairing) is sufficient to prevent any translocation through the Tat machinery.

Without a confirming the expression of these variants I cannot be certain whether these results are a result of a lack of expression or a genuine defect.

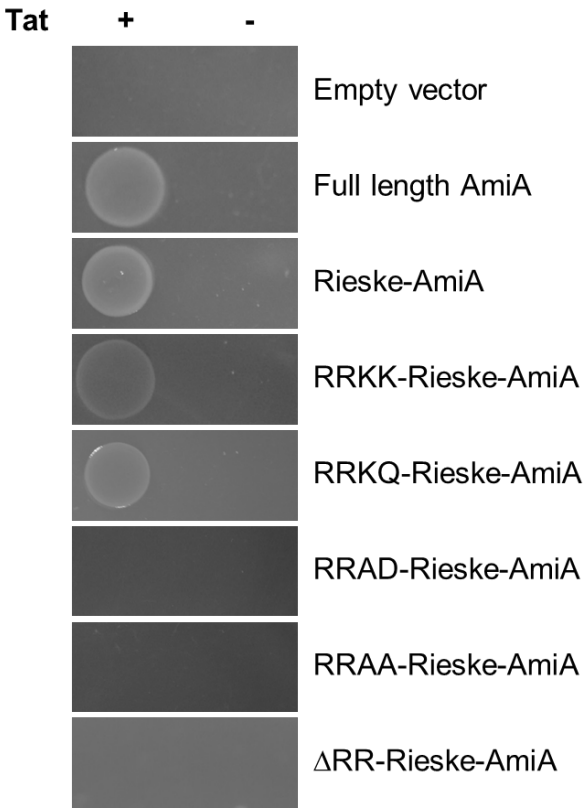


Figure 3.5 Assessing Tat-dependent translocation of Rieske-AmiA twin-arginine substitutions or deletions.
Spot tests of strains MCDSSAC (*tat*$^+$ strain) and MCDSSACΔtat (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), or the pSUPROM vector encoding full length AmiA, Rieske-AmiA or the twin-arginine variants RRKK-Rieske-AmiA, RRKQ-Rieske-AmiA, RRAD-Rieske-AmiA, RRAA-Rieske-AmiA and ΔRR-Rieske-AmiA. Strains were grown overnight in liquid media, subcultured to give an OD_{600} of 0.5 and 5 µl aliquots were spotted onto LB medium containing 2% (w/v) SDS. LB agar plates were incubated at 37°C for 16h.

3.2.2.5 Finding a reliable quantification assay for Tat-dependence using the Rieske-AmiA fusion

The SDS spot testing method used above can clearly indicate Tat-dependence of Rieske-AmiA fusion variants. However, ideally, a more quantitative assay is required where small growth differences between variants can be easily observed. To this end a serial dilution assay was optimised - the *tat*$^+$ and *tat*$^-$ strains expressing the twin-arginine variants of the Rieske-AmiA fusion were grown until they reached stationary phase in liquid media, diluted to give approximately 10, 100, 1,000 or 10,000 cells per 5µl aliquot and spotted onto standard (10g NaCl/l) or low salt (5g NaCl/l)-containing LB (as it was noted previously by Rebecca Keller that the level of salt in the LB may alter the behaviour

of the Rieske-AmiA fusion protein; personal communication). The plates also contained differing concentrations of SDS (1 or 2%).

As shown in Figure 3.6A, in the absence of SDS, all of the strains and plasmid combinations were able to grow equally well, indicating that none of the constructs were toxic to the cells. Figure 3.6 panels B & C indicate that when grown on low salt LB medium containing either 1 or 2% SDS variants of Rieske-AmiA that were previously shown in Figure 3.5 to be Tat-inactive (RRAD-Rieske-AmiA and RRAA-Rieske-AmiA) now allowed some growth of the *tat*$^+$ strain. In comparison, when the same strain and plasmid combinations were spotted onto standard LB medium containing 1 or 2% SDS the same variants were again unable to support any growth on SDS (Figure 3.6D & E). It is not clear why changing the level of salt should result in these differences, but it may relate to osmotic balance and cell wall stress.

Interestingly, it can be seen that although the RRKK-Rieske-AmiA still supports some growth of the *tat*$^+$ strain in standard LB medium containing 1% SDS, it was clearly much less than the 'wild type' twin-arginine-containing construct, and no growth at all was observed in the presence of 2% SDS in this experiment. Since the presence of 1% SDS allows a better differentiation between the severity of twin-arginine substitutions, it was decided that in all future experiments the assay for Tat-dependence of Rieske-AmiA will involve serial dilutions of stationary phase cultures and spotting onto standard LB medium containing 1% SDS, followed by growth at 37°C for 16 h.

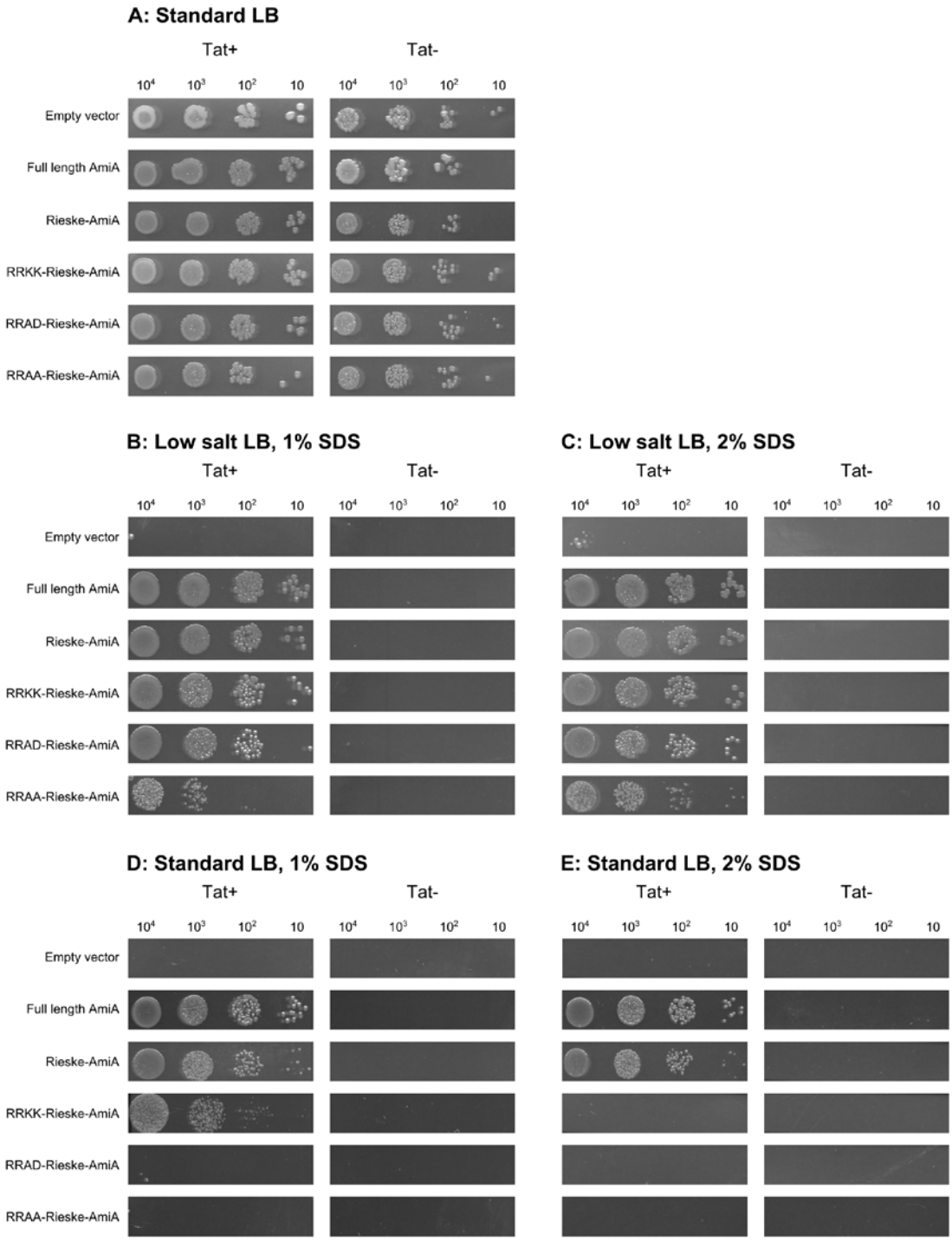


Figure 3.6 Tat-dependence of twin-arginine fusion variants on different compositions of LB agar plates containing 1-2% SDS.
Spot tests of strains MCDSSAC (*tat*$^+$ strain) and MCDSSACΔtat (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), or the pSUPROM vector encoding full length AmiA, Rieske-AmiA or the twin-arginine variants RRKK-Rieske-AmiA, RRAD-Rieske-AmiA and RRAA-Rieske-AmiA. Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto standard LB agar (panel A) or standard or low salt LB solid agar containing either 1% or 2% (w/v) SDS. LB agar plates were incubated at 37°C for 16 h. (A) standard LB medium (B) Low salt LB with 1% (w/v) SDS (C) Low salt LB with 2% (w/v) SDS (D) standard LB with 1% (w/v) SDS (E) standard LB with 2% (w/v) SDS.

3.2.3 Assessing Tat-dependence of the Rieske-Bla Reporter Fusion

3.2.3.1 Spot testing indicates partially Tat-dependent membrane insertion of Rieske-Bla

The fusion construct Rieske-Bla was tested for Tat-dependence using *tat*$^+$ and *tat*$^-$ *E. coli* strains, MC4100 and DADE, respectively. The *tat*$^-$ strain (DADE) lacks all known *tat* genes (Δ*tatABC*/Δ*tatE*) (Wexler *et al.*, 2000). The *tat*$^+$ and *tat*$^-$ strains harbouring Rieske-Bla, and a negative control of pSUPROM (empty vector) were plated and colonies were picked for dilution and spotting onto LB containing different concentrations of ampicillin.

If the Rieske-Bla fusion protein is fully Tat-dependent for the integration of TMD3, there should be no growth on ampicillin-containing media of the *tat*$^-$ strain producing this construct. It should be noted that unlike AmiA, Bla is a periplasmic protein that is normally exported co-translationally by the Sec pathway, therefore, it is a reporter that can use either Sec or Tat machinery for translocation depending on the signal sequence to which it is fused (Pradel *et al.*, 2009). Figure 3.7 shows that the Rieske-Bla fusion construct appears to be recognised by both of these pathways. The *tat*$^+$ strain has a higher level of ampicillin resistance than the *tat*$^-$ strain so there must clearly be some recognition and integration of the fusion protein by the Tat pathway (assuming that the level of the fusion protein is similar in the *tat*$^+$ and *tat*$^-$ strains). However the *tat*$^-$ strain, DADE, expressing the Rieske-Bla construct showed some growth on solid medium containing up to 37.5 µg/ml ampicillin. This result clearly shows that TMD3 of the Rieske protein is at least partly compatible with the Sec machinery. This is a new finding, as partial Sec recognition of TMD3 was not noted by Keller *et al*. (2012).

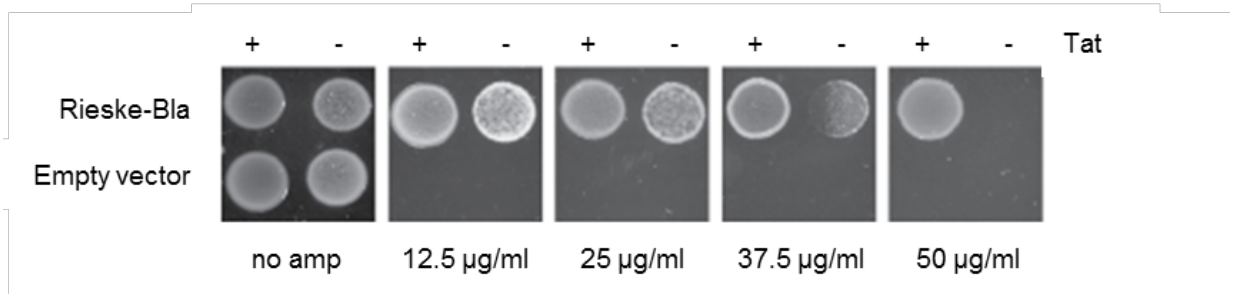


Figure 3.7 Translocation of the Rieske-Bla fusion protein is partially Tat-dependent.
Spot tests of strains MC4100 (*tat*$^+$ strain) and DADE (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), or the same vector encoding Rieske-Bla. An approximately equivalent number of colonies were resuspended in 100 µl of LB medium and 5 µl of the suspension was spotted onto LB medium containing ampicillin at concentrations of 12.5, 25, 37.5 and 50 µg/ml. The plates were incubated at 37°C for 16 h.

3.2.3.1.1 Testing Rieske-Bla twin-arginine variants for Tat-dependence

It was previously shown for the Rieske-AmiA reporter that introduction of the RRKK substitution unexpectedly still allowed some Tat-dependent transport, therefore, a similar twin-lysine substitution was also introduced into the Bla reporter to determine whether it would also still permit some Tat recognition. Unfortunately, it soon became clear that the spot test assay used in Figure 3.7 was rather variable, making it unreliable. To this end, a liquid growth assay in the presence of varying ampicillin concentrations was developed. Figure 3.8 shows growth of *tat*$^+$ and *tat*$^-$ strains harbouring Rieske-Bla and the RRKK, RRKQ and RRAD variants in liquid medium supplemented with varying concentrations of ampicillin. The behaviour of *tat*$^+$ cells producing the Rieske-Bla and the RRKK variant of Rieske-Bla appeared comparable, both showed similar growth kinetics in the presence of 12.5 and 25 µg/ml ampicillin, and showed some growth in the presence of 50 µg/ml ampicillin, as seen in the spot test (Figure 3.7). The RRKQ and RRAD variants of Rieske-Bla also supported some growth of the *tat*$^+$ strain in the presence of ampicillin, but not as well as the wild type or the RRKK variant. This indicates that in this assay these variants are to some extent Tat-dependent.

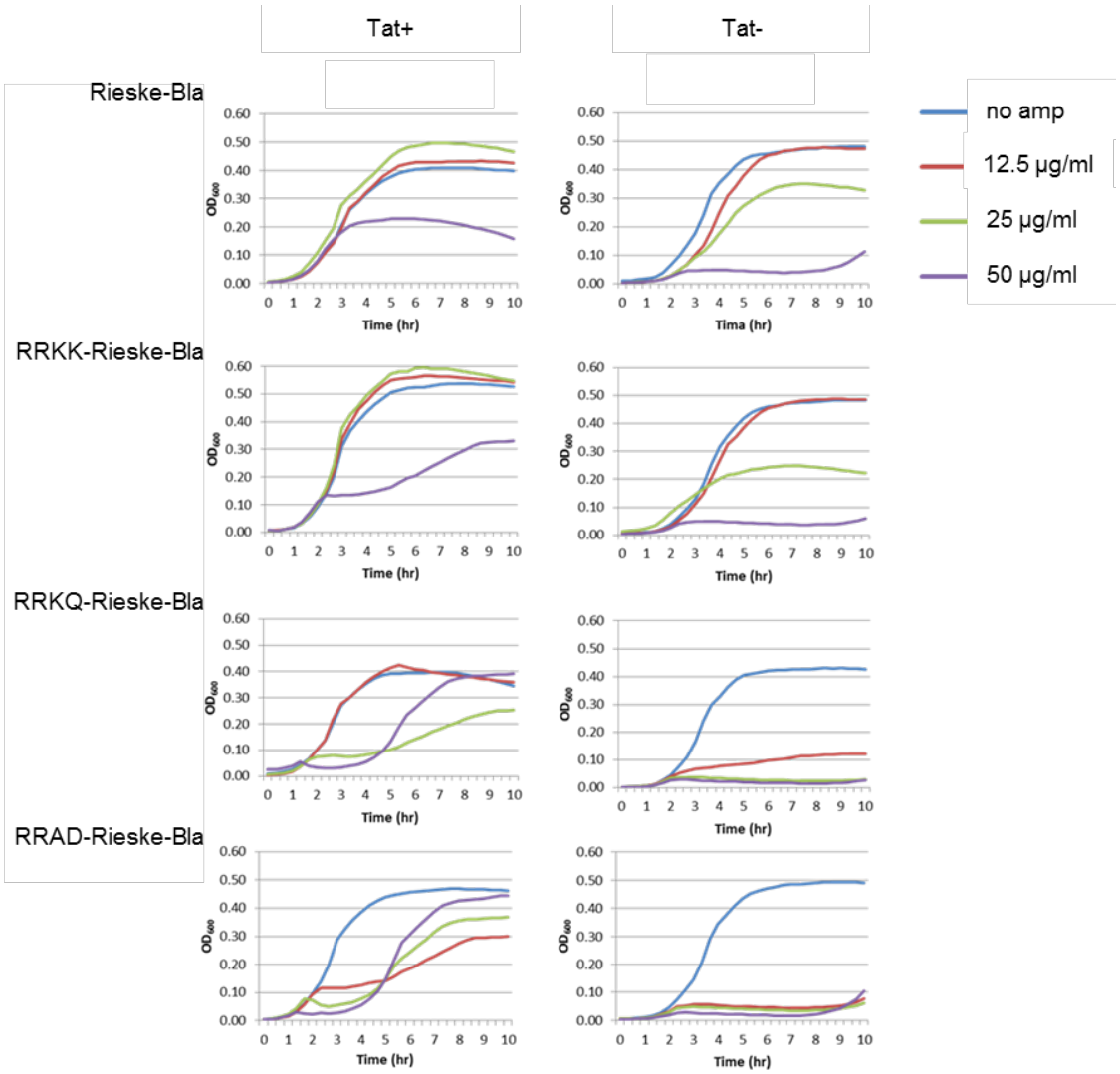


Figure 3.8 Tat- and Sec-dependent translocation of Rieske-Bla twin-arginine variants.
Representative growth curves of MC4100 (*tat*$^+$) and DADE (*tat*$^-$) strains harbouring pSUPROM vector encoding Rieske-Bla and twin-arginine variants RRKK-Rieske-Bla, RRKQ-Rieske-Bla and RRAD-Rieske-Bla. Stationary phase cultures were diluted to an OD_{600} of 0.03 with LB only or LB containing ampicillin at a final concentration of 12.5, 25 & 50 µg/ml ampicillin. Cultures were incubated for 10 h at 37°C with shaking and OD_{600} readings were taken every 20 min.

Comparison of growth seen for the *tat*$^-$ strain harbouring these fusion proteins yielded some interesting findings. It can be seen that the RRKQ and RRAD variants of Rieske-Bla produced in the *tat*$^-$ strain supported no significant growth when the medium contained ampicillin. This would suggest that these two substitutions result in TMD3 no longer being recognised by the Sec pathway (but must still allow some recognition by the Tat pathway to account for the growth seen in the *tat*$^+$ strain). By contrast, the Rieske-Bla and RRKK-Rieske-Bla constructs supported some growth at 25 µg/ml

ampicillin, and so TMD3 must still be partially recognised by the Sec pathway in these cases.

To confirm that the lack of growth in ampicillin-containing media for the *tat*$^-$ strain producing the RRKQ-Rieske-Bla and RRAD-Rieske-Bla fusions is due to an effect of the amino acid substitutions and not to lack of protein production or protein stability, whole cells of the *tat*$^+$ and *tat*$^-$ strains producing these fusion constructs were analysed by Western blotting using a polyclonal anti-Bla antibody. Figure 3.9 shows that the Rieske-Bla variant proteins are produced in both *tat*$^+$ and *tat*$^-$ strains, although the levels of protein detected are quite variable. Nonetheless, since RRKQ-Rieske-Bla and RRAD-Rieske-Bla can be detected in the *tat*$^-$ strain, the absence of growth must be due to effects of the substitutions on recognition by the Sec pathway.

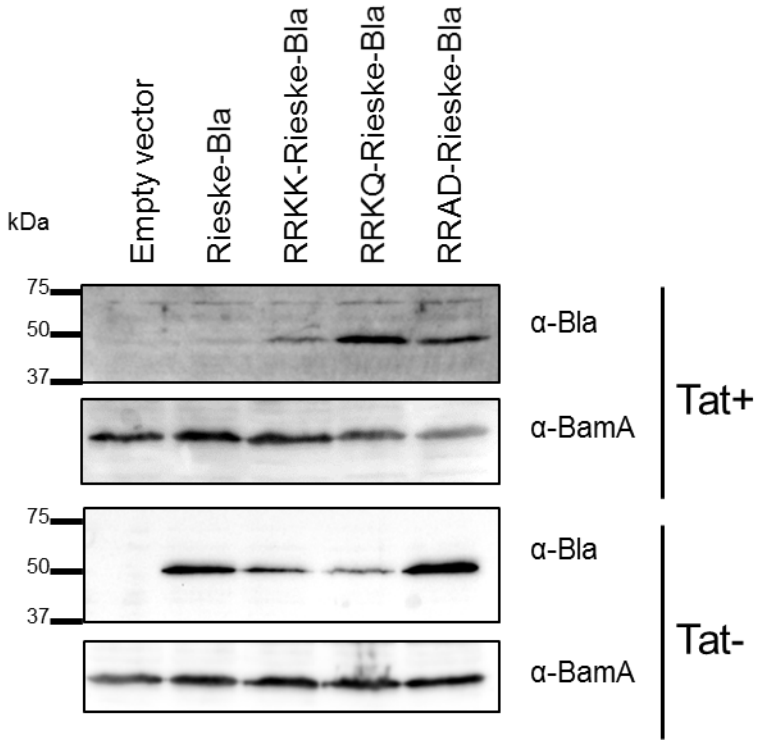


Figure 3.9 The Rieske-Bla twin-arginine variants are stably produced.
Whole cells of MC4100 (*tat*$^+$) and DADE (*tat*$^-$) strains harbouring an empty vector (pSUPROM) or the pSUPROM vector encoding Rieske-Bla or variants RRKK-Rieske-Bla, RRKQ-Rieske-Bla and RRAD-Rieske-Bla were separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Bla antibody. As a loading control the levels of the OM protein BamA were detected using an anti-BamA antiserum.

Analysis of the effect of twin-arginine mutations on the membrane integration of the Rieske-Bla construct is complicated by the fact that the translocation of Bla is mediated by both Sec and Tat. Some of the substitutions (for example RRAD) clearly abolished recognition of TMD3 by the Sec pathway as the *tat*$^-$ strain harbouring this variant was unable to grow even at the lowest level of ampicillin used. This is not surprising as negative charges present in the n-regions of Sec signal peptides are known to retard translocation by the Sec pathway (Vlasuk *et al.*, 1983, Inouye *et al.*, 1982).

The fact that the Rieske-Bla fusion shows dual recognition of TMD3 means that it will not be a particularly useful tool to examine the exact requirements for recognition by the Tat pathway, as observations will be complicated by underlying Sec recognition (which will also vary when site-directed substitutions are introduced). However, Rieske-Bla is a useful construct in the selection of mutations that enhance interaction with the Sec pathway, since this should lead to growth on high concentrations of ampicillin in *tat* mutant strains. For the rest of this project Rieske-Bla will be used to analyse the insertion of Rieske by the Sec machinery and all tests for Sec-dependence will be undertaken using only the *tat*$^-$ strain (DADE; Δ*tatABC*/Δ*tatE*) *as* this shows only Sec-translocation of Rieske-Bla.

3.2.3.2 Finding an assay to reliably show Sec translocation of the Rieske-Bla fusion

A reliable, robust and quantitative assay needed to be found to analyse the extent of Sec translocation of each of the Rieske-Bla fusion variants. To develop this, Oxoid M.I.C.Evaluator™ (Thermo Fisher Scientific) test strips were used. These contain a continuous antibiotic concentration gradient, in this case ampicillin, with a gradient from 0-256 µg/ml. The strip is placed on an LB agar plate containing a lawn of bacteria and after incubation, a symmetrical inhibition ellipse centred along the strip is seen. The minimum inhibitory concentration (M.I.C.) value (in µg/ml) is read from the scale where the pointed end of the ellipse intersects the strip.

3.2.3.2.1 Minimum Inhibitory Concentration Evaluator test strips confirms Sec and Tat translocation of the Rieske-Bla fusion

Firstly the M.I.C.Evaluator™ test strip was tested against the necessary controls; cells only of both the *tat*$^+$ and *tat*$^-$ *s*trains and the same strains harbouring the empty vector (pSUPROM) or the pSUPROM vector containing Rieske-Bla. As Figure 3.10 indicates, the M.I.C. for *E. coli* alone, in the absence of any plasmid is 1-2 µg/ml and as expected the introduction of the empty pSUPROM vector (which specifies kanamycin resistance) in these cells does not change this M.I.C. When the Rieske-Bla variant is expressed in the *tat*$^+$ strain and tested, the M.I.C. increases to 16 µg/ml. This is due to the translocation of the Bla into the periplasm *via* both the Sec and the Tat pathways. In comparison when this variant is expressed in the *tat*$^-$ *s*train the M.I.C. is 8 µg/ml which is a measure of the level of Bla translocated by the Sec machinery. This would appear to be a reliable and quantifiable assay that could support investigation into the Sec recognition of Rieske TMD3. Therefore, this will be the standard assay for further work using Rieske-Bla.

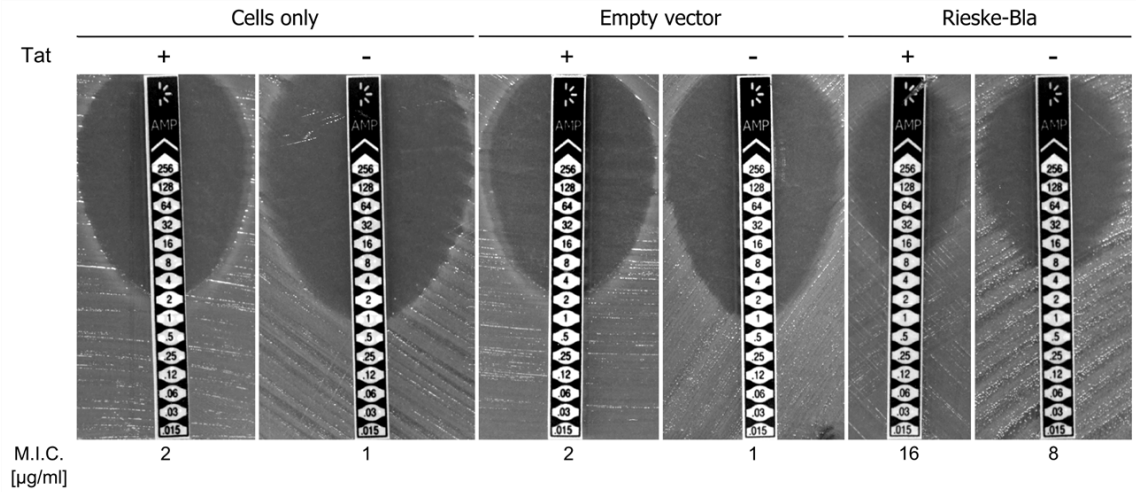


Figure 3.10 M.I.C.Evaluator™ test strips show Sec and Tat translocation of the Rieske-Bla fusion. Representative images of M.I.C.Evaluator™ strip tests of strains MC4100 (*tat*$^+$) and DADE (*tat*$^-$) alone (cells only), harbouring the pSUPROM plasmid (Empty vector) or pSUPROM encoding Rieske-Bla. Stationary phase cultures were diluted to OD_{600} of 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plates were incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip by the clearing of bacteria. The average M.I.C for these variants is located at the bottom of the test strip.

3.2.3.2.2 Testing Rieske-Bla twin-arginine variants for Sec translocation

The first Rieske-Bla variants tested using the M.I.C.Evaluator™ assay were the twin-arginine variants that were generated above and previously tested in liquid media, allowing comparison of the two testing methods. As with the previous results using the M.I.C.Evaluator™ assay, Figure 3.11 shows that when the *tat*⁻ *s*train produces some of the twin-arginine variants there is detectable translocation through the Sec machinery. Figure 3.11A shows the results from the M.I.C.Evaluator™ test strip analysis of the *tat*⁻ *s*train producing Rieske-Bla, RRKK-Rieske-Bla, RRAD-Rieske-Bla, RRAA-Rieske-Bla, RRKQ-Rieske-Bla and ΔRR-Rieske-Bla and Figure 3.11B is the analysis of triplicate results showing the mean M.I.C. and standard deviation (s.d.) for the *tat*⁻ *s*train producing each Rieske-Bla variant.

Strain DADE (*tat*⁻) producing ΔRR-Rieske-Bla has a similar M.I.C. to DADE producing Rieske-Bla, indicating that the twin-arginines are not essential for Sec translocation of TMD3. The RRAA-Rieske-Bla fusion variant specified a slightly decreased M.I.C. whereas both RRKK-Rieske-Bla and RRKQ-Rieske-Bla yielded significantly decreased M.I.C.s, however, they still showed some translocation through the Sec machinery. This is in contrast to RRAD-Rieske-Bla, which as previously shown (Figure 3.8) prevents translocation through the Sec machinery with an M.I.C. the same as cells only (Figure 3.10, 1 µg/ml) due to the negative charge replacement of the twin-arginine. These results corroborate previously shown data for Rieske-Bla variants and demonstrate the reliability of this test.

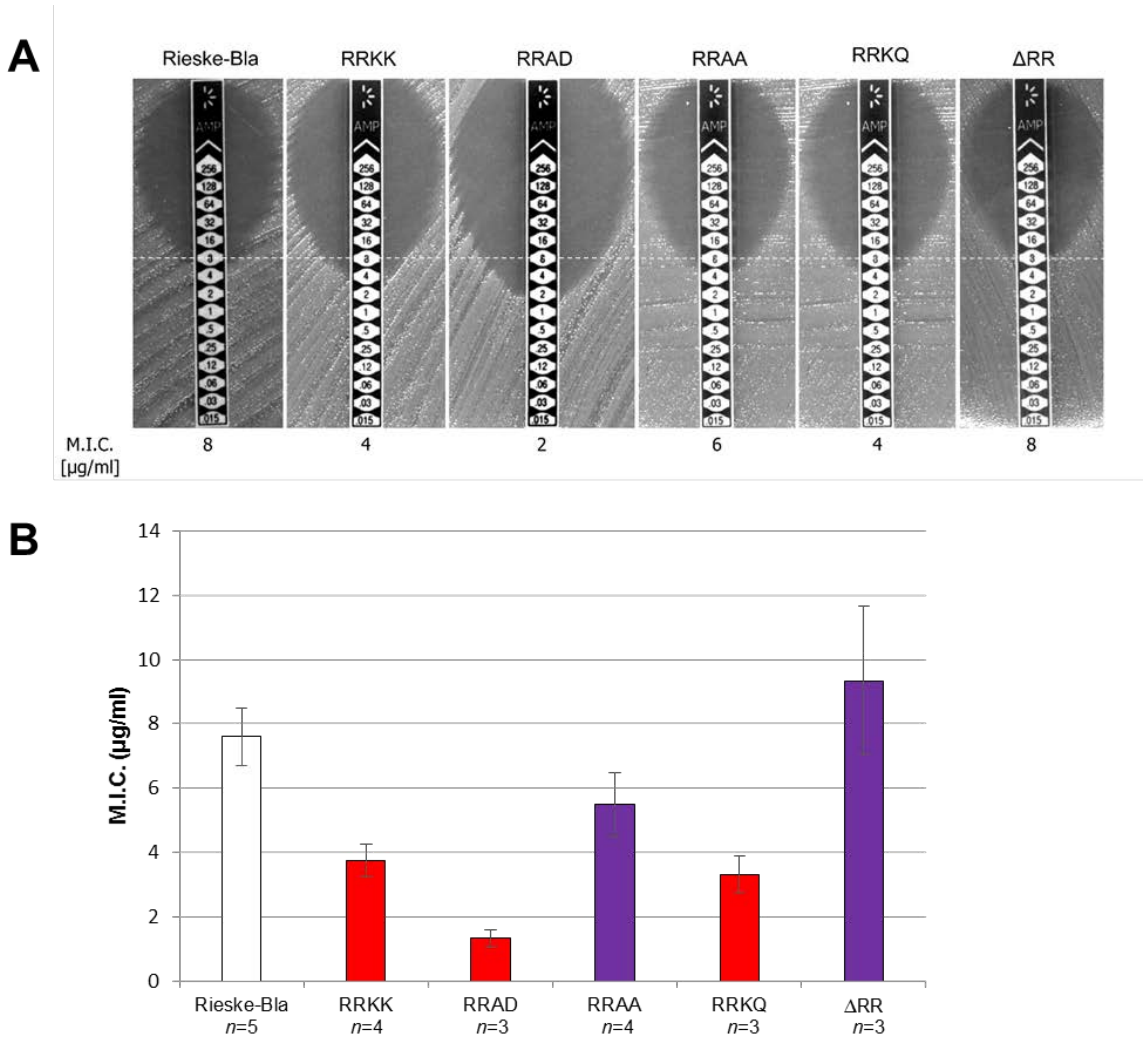


Figure 3.11 Sec-dependent translocation of Rieske-Bla twin-arginine variants.
(A) Representative M.I.C.Evaluator™ strip test of strain DADE (*tat*⁻) harbouring the pSUPROM vector encoding Rieske-Bla or twin-arginine variants RRKK-Rieske-Bla, RRAD-Rieske-Bla, RRAA-Rieske-Bla, RRKQ-Rieske-Bla and ΔRR-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} of 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of Rieske-Bla (8 µg/ml).
(B) The mean of the M.I.C.Evaluator™ assay data with the s.d. (error bars) calculated and plotted on a graph. The number of biological replicates is indicated under the graph. Significance is calculated by the Mann-Whitney Rank Sum Test and if $p<0.04$ then the bars are red, representing a significant decrease in M.I.C. compared to Rieske-Bla, purple shows no significance calculated.

3.3 Discussion

S. coelicolor Rieske is a dual-targeted cytoplasmic membrane protein that is inserted into the membrane using both Sec and Tat translocation pathways. While this dual-targeting of Rieske has been investigated in its native *S. coelicolor* (Hopkins *et al.*, 2014) and within *E. coli* (Keller *et al*., 2012), the specific requirements of this insertion process are not known.

Exploring insertion of the membrane portion of Rieske required the use of reliable fusion reporter proteins. These are tools that are used to probe transport pathways and determine transmembrane protein topologies. A good phenotypic screening assay was required to assess the output by the reporter fusion protein. Keller *et al.* (2012) used two fusion reporters, MBP and AmiA, to assess insertion of Rieske by Sec and Tat. Although these reporters were of sufficient utility to assess the insertion of Rieske into the cytoplasmic membrane, MBP is not a reliable enough reporter to allow further investigation into the dual targeting mechanism. AmiA is a reliable reporter but the phenotypic screening assay needed to be refined to assess small differences in growth supported by the Rieske-AmiA variants.

A consistent and reliable assay was developed using the fusion reporter AmiA which enabled discrimination between relatively small differences in Tat translocation caused by mutations in Rieske. For Tat substrates, the twin-arginines of the signal peptide are essential for recognition by the Tat machinery. Substitution of these arginines for twin-lysine has been found to abolish Tat translocation for many substrates (Bachmann *et al.*, 2006, Buchanan *et al.*, 2001, Widdick *et al.*, 2008, Stanley *et al.*, 2000). However, when produced from the vector pSUPROM, a twin-lysine substitution of Rieske-AmiA variant was still able to translocate AmiA to the periplasm through the Tat pathway. Interestingly, this is in contrast to previous work of Keller *et al.* (2012), and further analysis revealed that the expression level of the fusion protein has a major impact on the translocation to the periplasm.

The second reporter developed in this study was Bla. Bla is a well-known reporter for topology studies and has been shown to be compatible with both Sec and Tat pathways for transport into the periplasm (Stanley *et al.*, 2002, McCann *et al.*, 2007). In this case, Bla is a robust Sec reporter when expressed in the *tat*$^{-}$ strain and a reliable assay has been developed based on ampicillin-impregnated test strips that is able to quantify small differences in Sec translocation specified by Rieske-Bla variants.

In conclusion, in this Chapter, two reliable Rieske fusion reporter proteins have been developed and robust assay systems to assess the insertion and translocation of Rieske-AmiA and Rieske-Bla variants by the Tat and Sec machineries have been established. These assays will be used in future Chapters to investigate the co-operation of Sec and Tat translocation machineries in the insertion of *S. coelicolor* Rieske into the *E. coli* cytoplasmic membrane.

4 Sec translocation of *Streptomyces coelicolor* Rieske-Bla

4.1 Introduction

4.1.1 Topological determinants

As discussed in Chapter 1, the insertion propensity of a TMD into the cytoplasmic membrane by the Sec machinery depends on a number of determinants within the helix; overall hydrophobicity, length and location of charged and polar residues (Hessa *et al.*, 2005, Hessa *et al.*, 2007). Other factors are also involved, including the nature of residues present in internal and external loop regions, and the thermodynamics of protein-lipid interactions (Hessa *et al.*, 2005). Figure 4.1 illustrates the most important of these topological determinants; hydrophobicity, charge difference across the membrane and the positive inside rule.

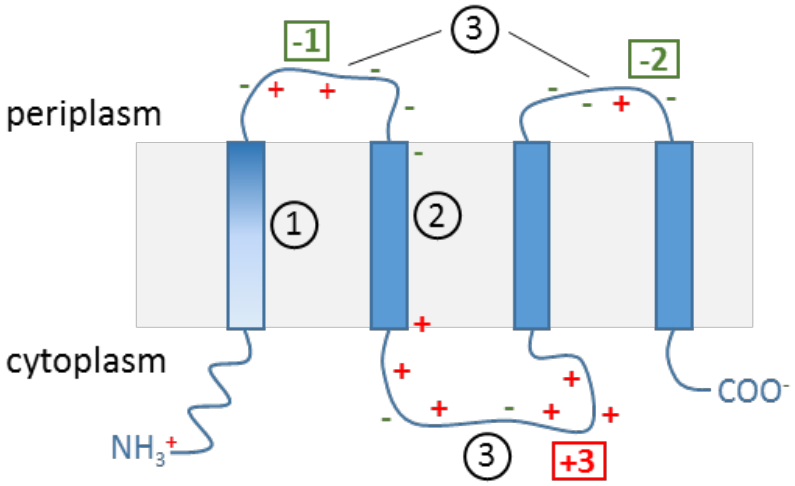


Figure 4.1 The main topological determinants for the Sec machinery to correctly insert a TMD.
1) Hydrophobicity is the driving force for TMD insertion into the membrane 2) The charge difference within a TMD is based on the location of charged residues within the helix 3) The positive inside rule specifies that positive charges are more likely to be located in the cytoplasmic loop regions.

4.1.2 Hydrophobicity

Overall hydrophobicity of a TMD is the driving force for insertion into the cytoplasmic membrane from the Sec machinery; a stretch of hydrophobic residues cause the opening of the lateral gate to allow the release of the helix into the membrane (von Heijne, 1997).

Typically the hydrophobicity of a TMD is additive, the more hydrophobic residues within the helix the greater the insertion propensity (White & von Heijne, 2008).

Biological hydrophobicity scales have been produced that measure the apparent free energy of insertion of a TMD by the Sec61 complex into the ER membrane using an *in vitro* system (Hessa *et al.*, 2005) and this has been extended experimentally to establish a position-specific biological hydrophobicity scale (Hessa *et al.*, 2007). The methodology for these experiments makes use of an engineered version of the *E. coli* IM protein Leader peptidase, Lep, which consists of two TMD which insert into the membrane with both N- and C-terminus in the lumen, followed by a large luminal P2 domain. A hydrophobic segment (H-segment) to be tested is engineered into the normally luminal P2 domain of Lep that is flanked at both sides by acceptor sites for N-linked glycosylation. This engineered Lep protein is expressed in ER-derived dog pancreas rough microsomes, the degree of integration of the H-segment is the difference between a singly or doubly glycosylated protein, which is quantified by SDS-PAGE and Western blotting. This is then converted into the apparent free energy of insertion of the H-segment (ΔG_{app}).

Each individual amino acid was placed in the centre of a 19 residue H-segment and the free energy of insertion was calculated to generate a 'biological' scale (Hessa *et al.*, 2005). The position specificity of every amino acid was also investigated by systematically scanning each residue across an H-segment to calculate the effect of each residue on the overall insertion into the membrane (Hessa *et al.*, 2007). Both of these scales generated a full quantitative explanation of TMD recognition by the Sec61 translocon for insertion into the ER membrane.

These scales have subsequently been investigated for the SecYEG-mediated insertion of TMDs into the bacterial IM (Ojemalm *et al.*, 2013). The methodology used here differs as in this case two hydrophobic segments were inserted into P2 of Lep; the engineered H-segment, as previously described, and a reporter TMD (R-segment) identical to TMD2 of the *E. coli* IM protein lactose permease (LacY). This R-segment is cleaved by the

intramembrane rhomboid protease, GlpG, if inserted into the membrane with an N_{out}-C_{in} orientation. Therefore, when an H-segment is efficiently inserted into the membrane there is no cleavage as the R-segment is inserted in N_{in}-C_{out} orientation producing a full-length form or Lep^{LacY}. On the other hand, if the H-segment is unable to insert into the membrane, the R-segment is incorrectly inserted, leading to cleavage by GlpG and yielding a truncated form of Lep^{LacY}. The comparison between these forms can determine the degree of insertion of the H-segment. Using similar experimental engineering of residues within the H-segment as Hessa *et al.* (2005; 2007), the ΔG_{app} was calculated for each individual amino acid and how its position in the TMD affects the insertion of the H-segment into the bacterial IM. It was found that the positional relevance of residues in the TMD overall correlates well with previous values for the mammalian ER (Hessa *et al.*, 2007, Hessa *et al.*, 2005). However, the hydrophobicity threshold was significantly lower than that measured for TMD insertion by Sec61 into the mammalian ER membrane and a little higher than the value for a YidC-mediated TMD insertion into the *E. coli* IM (Xie *et al.*, 2007).

A hydrophobicity threshold needs to be overcome before insertion of a hydrophobic stretch is possible. Marginally hydrophobic TMDs are unable to reach this threshold without assistance from neighbouring TMDs and charged residues flanking the TMD. Both the influence of other TMDs and positive charges can contribute to lowering the hydrophobicity threshold for this type of TMD (Ojemalm *et al.*, 2012, Lundin *et al.*, 2008).

4.1.3 Positive inside rule

Charges have been found to be critical in topology decisions by the Sec machinery. This is particularly notable within cytoplasmic loop regions where arginine and lysine residues are found to be four times more prevalent compared to periplasmic loops (Heijne, 1986). This is known as the positive inside rule. Positive charges are more likely to be found in the cytoplasmic loop regions due to the energetic cost for the Sec machinery to

translocate them across the membrane into the periplasm against the proton gradient. This positive inside rule appears to be applied universally to IM proteins that are found in the ER, plasma membrane, inner mitochondrial membrane, chloroplast thylakoid membrane and the bacterial IM (Gavel *et al.*, 1991, von Heijne & Gavel, 1988).

The positive inside rule has been fully explored in a small multidrug transporter, EmrE, which has inherent properties allowing it to adopt two different orientations in the membrane (Rapp *et al.*, 2007). Its lack of positive charges leads to a small charge bias between internal and external loops. Therefore, it is possible to fix the topology of this protein by the addition of a positive charge in one of these loops (Rapp *et al.*, 2006) and a single positive charge located in the very C-terminus can cause a topology inversion (Seppala *et al.*, 2010). Manipulation of the location of positive charges within the loop regions can also invert the topology of another protein, Lep, containing two TMDs (Nilsson & von Heijne, 1990, von Heijne, 1989). Positive charges have also been found to affect the insertion of the neighbouring TMD (Lerch-Bader *et al.*, 2008, Hessa *et al.*, 2007) indicating that topological determinants can have an additive effect on the Sec machinery for correct insertion of TMDs.

4.2 Aims

It was seen in Chapter 3 that there was unexpectedly some interaction of the Sec machinery with TMD3 of Rieske, which was revealed when Bla was used as the fusion reporter. The aim of this Chapter was to further assess the interaction of the Sec machinery with Rieske, to dissect whether there are features within the TMD region of Rieske that modulate interactions with the Sec machinery, and in particular to promote the release of TMD3. To do this a full programme of mutagenesis was undertaken with the Rieske-Bla fusion and variants generated were tested for translocation by determining the M.I.C. of ampicillin. Variants of Rieske-Bla that differed significantly from the ‘wild-type’ in terms of the level of ampicillin resistance they conferred were further analysed biochemically to support the results of the phenotypic observations.

4.3 Results

4.3.1 Investigating the cytoplasmic loop between TMD2 and TMD3

Keller *et al.* (2012) undertook a series of TMD-swapping experiments, where TMD1 and TMD2 of Rieske were replaced by the first two TMDs of a typical SRP-targeted Sec-dependent membrane protein, MtlA (Beck *et al.*, 2000). Furthermore, Keller *et al*. (2012) also increased the length and hydrophobicity of the Rieske transmembrane region by fusing these same two TMDs of MtlA N-terminal to Rieske's TMD1 and TMD2, to give a five TMD protein. In both of these cases, the insertion of the last TMD of Rieske remained Tat dependent, thereby indicating that there is nothing unusual or special about TMD1 and TMD2 of Rieske that facilitates the dual recognition by Sec and Tat. Therefore it was concluded that the cytoplasmic loop region between TMD2 and TMD3 must be critical to facilitate the co-operation of the Sec and Tat machineries, acting as a stop transfer signal for the Sec machinery and allowing recognition of a tethered twin-arginine signal sequence by the Tat machinery.

There is clear sequence conservation present in this loop region between different actinobacterial Rieske proteins, as shown in Figure 4.2. The linker length, between the end of TMD2 and the twin-arginine of the Tat motif, is always 43 amino acids long. Moreover, the loop contains a cluster of 4 negative amino acids, a highly conserved RH motif and a predicted α-helix region. The Tat motif is located N-terminal to TMD3 and as was seen in Chapter 3, it is essential for recognition by the Tat machinery. As these features appear to be conserved across the different Rieske proteins, they were deemed important for the release of the polypeptide from the Sec machinery and/or the recognition of the tethered signal sequence by the Tat machinery. In this Chapter the translocation of Rieske-Bla by the Sec machinery was investigated.

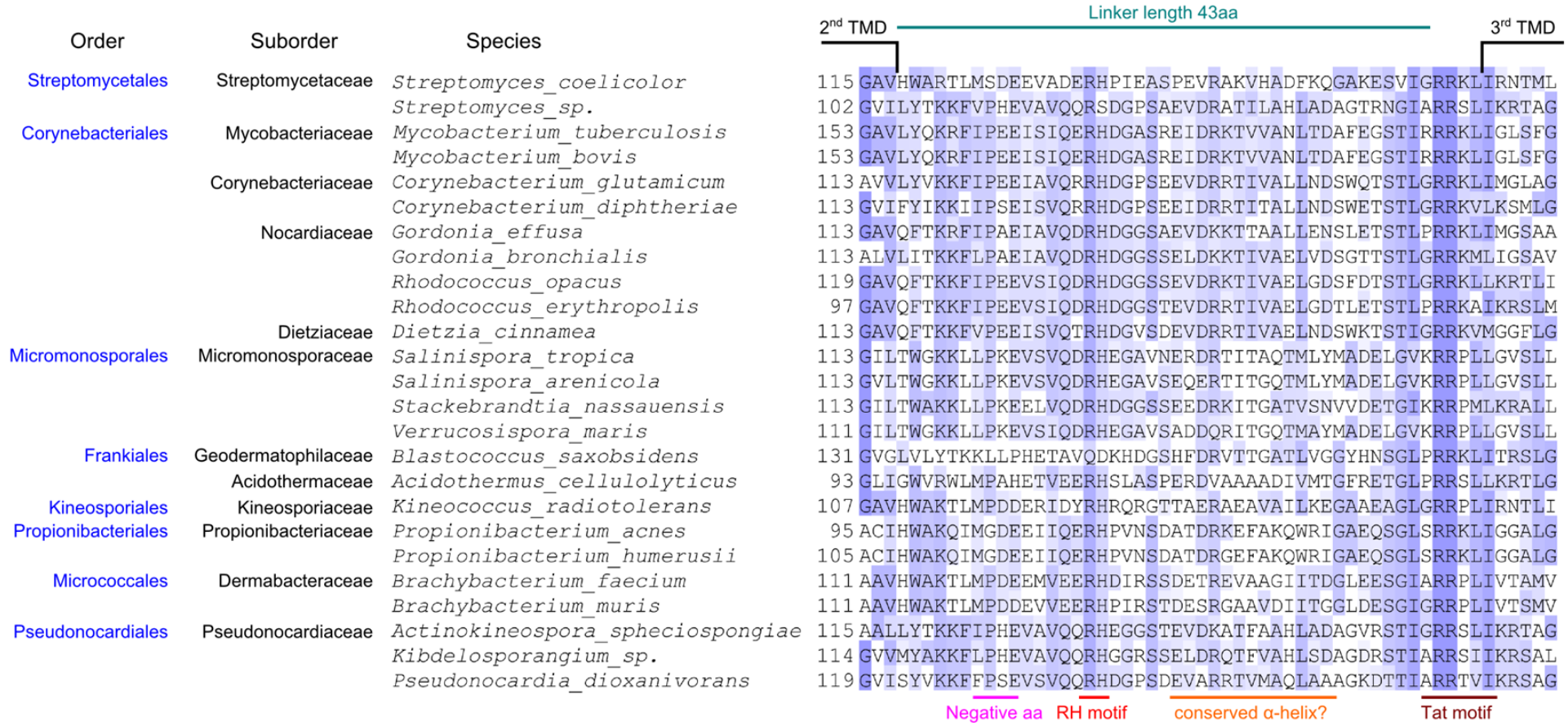


Figure 4.2 Alignment of Actinobacterial Rieske proteins loop region between TMD 2 and TMD 3
Protein sequence alignment of the Rieske FeS proteins for different Actinobacteria (as indicated on the left – order, suborder and species). Proteins were aligned using Clustal WS and Jalview (Waterhouse *et al.*, 2009). Conserved regions are indicated under the alignment; negative amino acids (pink), RH motif (red) potential conserved α-helix (orange) and the Tat motif (brown). Transmembrane domains (TMD) were predicted using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/).

4.3.1.1 Mutations generated in the cytoplasmic loop region have little effect on the insertion of TMD3 by the Sec machinery

The first step in understanding whether the conserved residues in the cytoplasmic loop region are important for modulating interaction with the Sec pathway was to undertake a systematic site-directed mutagenesis programme where each of the conserved residues in the loop region encoded by the pSUPROM Rieske-Bla construct were individually mutated. The cytoplasmic loop region is shown above Figure 4.4 and conserved residues that were mutated in this Chapter are highlighted in colour. These variants were produced in DADE (*tat*$^-$ strain) to allow Sec-dependence to be scored, and tested for the M.I.C. of ampicillin using M.I.C.Evaluator^{TM} test strips, as undertaken previously in Chapter 3 and a representative example is seen in Figure 4.3, the raw data can be found in the Appendix. Experiments were performed in triplicate and the mean of these data was assessed for significance compared to Rieske-Bla using Mann-Whitney Rank Sum Test. Figure 4.4 and Table 4.1 show the results of these site-directed mutagenesis experiments, alongside the M.I.C. determined for the twin-arginine variants that were previously described in Chapter 3.

4.3.1.1.1 Conserved amino acid substitutions have little effect on Sec translocation

The first set of residues that were mutated in this study was the RH motif (R133 and H134, coloured green in the sequence above Figure 4.4) as apart from the invariant twin-arginine motif this is the most conserved pair of residues present in the loop region. Two substitutions of these residues were made: R133H H134R (RH-HR) and R133K H134K (RR-KK). As shown in Figure 4.3 these substitutions did not significantly change the M.I.C. for ampicillin compared to ‘wild-type’ Rieske-Bla. It can therefore be concluded that these mutations do not change Sec translocation of TMD3.

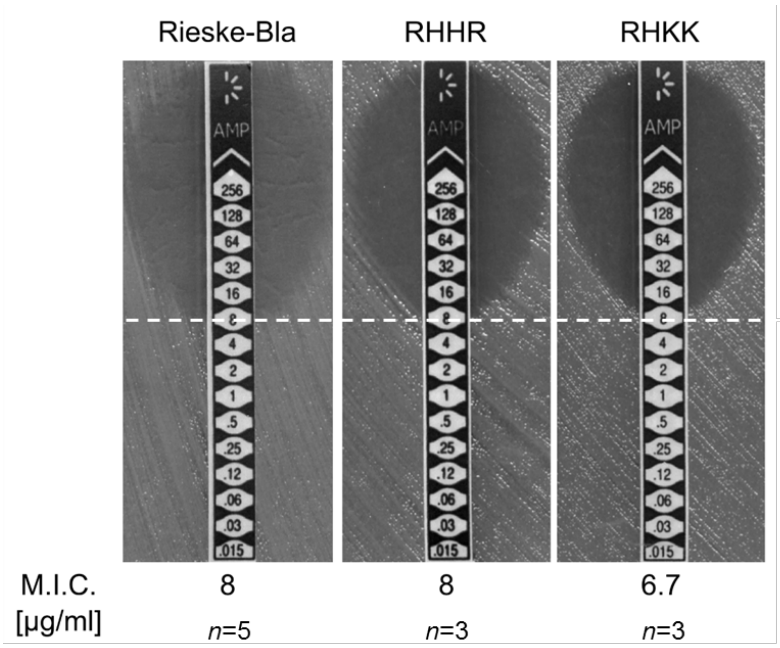


Figure 4.3 RH motif mutations has no effect on Sec translocation.
Representative M.I.C.Evaluator™ strip test of DADE (*tat*⁻) harbouring the pSUPROM vector encoding Rieske-Bla or RH motif variants RHHR-Rieske-Bla and RHKK-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of Rieske-Bla (8 µg/ml). The mean M.I.C. value and the number of biological replicates for these variants is given below the images.

Figure 4.4 Sec dependent translocation of Rieske-Bla fusion variants.
M.I.C.Evaluator™ strip test of strain DADE (*tat*⁻) harbouring the pSUPROM vector encoding the Rieske-Bla fusion variants indicated on the graph. Stationary phase cultures were diluted to OD_{600} of 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The mean of the M.I.C.Evaluator™ assay data with the s.d. (error bars) calculated (as seen in Table 4.1) and plotted on the graph. The number of biological replicates for each variant can be found in Table 4.1. Significance compared to Rieske-Bla is calculated by the Mann-Whitney Rank Sum Test and if $p<0.04$ then the bars are green (significant increase in M.I.C.) or red (significant decrease in M.I.C.). Above the graph is the conserved loop region for *S. coelicolor* Rieske where the conserved residues that were mutated are indicated in colour.

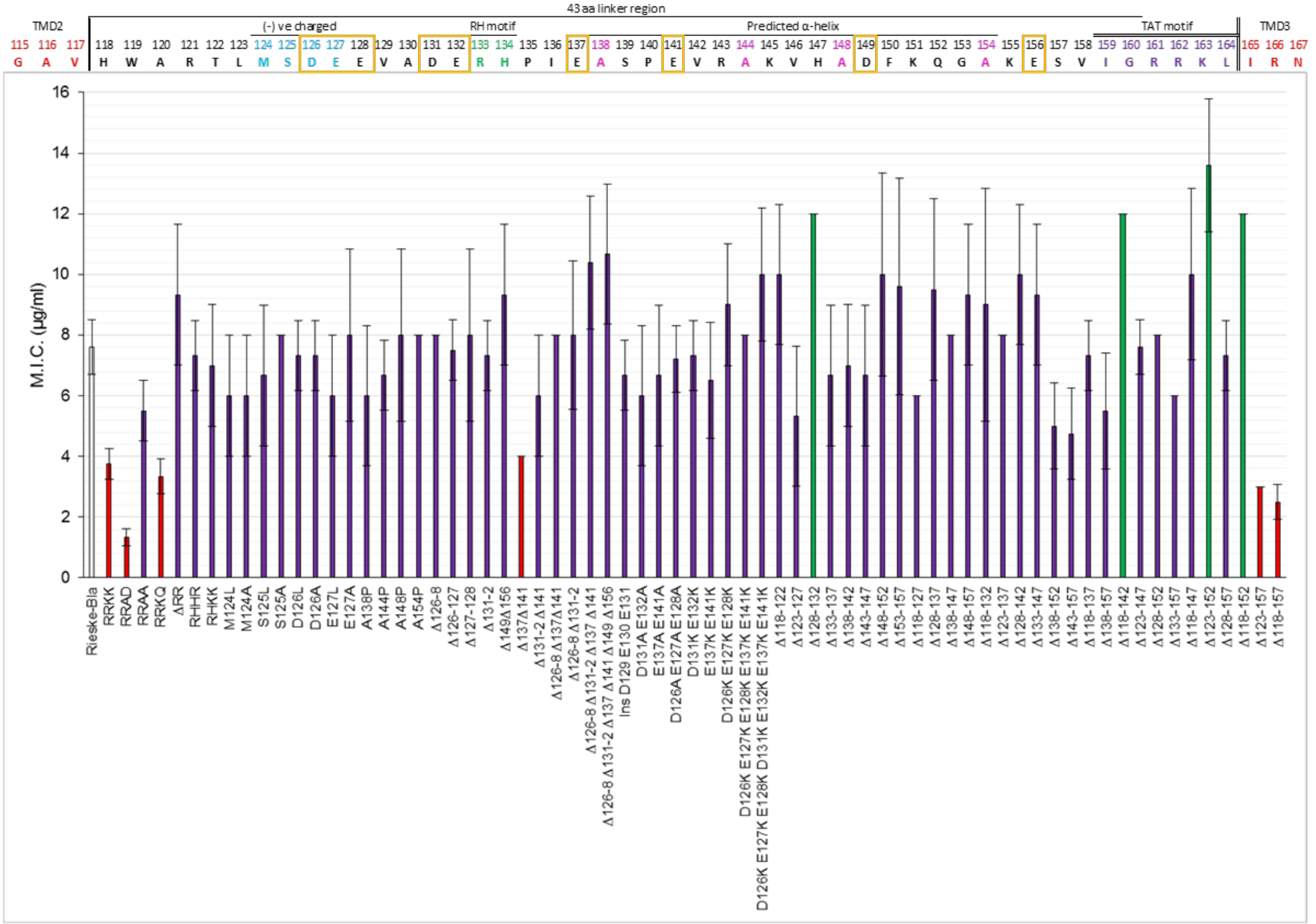

43 aa linker region
TMD2
(-) ve charged
RH motif
Predicted α-helix
TAT motif
TMD3
115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167
G A V H W A R T L M S D E E V A D E R H P I E A S P E V R A K V H A D F K Q G A K E S V I G R R K L I R N
M.I.C. (µg/ml)
0 2 4 6 8 10 12 14 16
Rieske-Bla
RRKK
RRAD
RRAA
RRKQ
ΔRR
RHHR
RHKK
M124L
M124A
S125L
S125A
D126L
D126A
E127L
E127A
A138P
A144P
A148P
A154P
Δ126-8
Δ126-127
Δ127-128
Δ131-2
Δ149Δ156
Δ137Δ141
Δ131-2 Δ141
Δ126-8 Δ137Δ141
Δ126-8 Δ131-2
Δ126-8 Δ131-2 Δ137 Δ141
Δ126-8 Δ131-2 Δ137 Δ141 Δ149 Δ156
Ins D129 E130 E131
D131A E132A
E137A E141A
D126A E127A E128A
D131K E132K
E137K E141K
D126K E127K E128K
D126K E127K E128K E137K E141K
D126K E127K E128K D131K E132K E137K E141K
Δ118-122
Δ123-127
Δ128-132
Δ133-137
Δ138-142
Δ143-147
Δ148-152
Δ153-157
Δ118-127
Δ128-137
Δ138-147
Δ148-157
Δ118-132
Δ123-137
Δ128-142
Δ133-147
Δ138-152
Δ143-157
Δ118-137
Δ138-157
Δ118-142
Δ123-147
Δ128-152
Δ133-157
Δ118-147
Δ123-152
Δ128-157
Δ118-152
Δ123-157
Δ118-157

Table 4.1 Sec dependent translocation of Rieske-Bla fusion variants
M.I.C.Evaluator™ strip test of strain DADE (*tat*⁻) harbouring the pSUPROM vector encoding the Rieske-Bla fusion variants indicated in the table. Stationary phase cultures were diluted to OD_{600} of 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The mean of the M.I.C.Evaluator™ assay data with the s.d. and number of biological replicates, n (in brackets) is shown in the table. Significance compared to Rieske-Bla is calculated by the Mann-Whitney Rank Sum Test and if $p<0.04$ then the numbers are shown in green (significant increase in M.I.C.) or red (significant decrease in M.I.C.).

	Variant	Average M.I.C. (s.d.)		Variant	Average M.I.C. (s.d.)
	Rieske-Bla	7.6 (0.9,n=5)	Truncations	Δ118-122	10.0 (2.3,n=4)
TAT motif	RRKK	3.8 (0.5,n=4)		Δ123-127	5.3 (2.3,n=4)
	RRAD	1.3 (0.3,n=3)		Δ128-132	12.0 (0.0,n=4)
	RRAA	5.5 (1.0,n=3)		Δ133-137	6.7 (2.3,n=3)
	RRKQ	3.3 (0.6,n=3)		Δ138-142	7.0 (2.0,n=4)
	ΔRR	9.3 (2.3,n=3)		Δ143-147	6.7 (2.3,n=3)
RH motif	RHHR	7.3 (1.2,n=3)		Δ148-152	10.0 (3.3,n=6)
	RHKK	7.0 (2.0,n=3)		Δ153-157	9.6 (3.6,n=5)
Negatively charged	M124L	6.0 (2.0,n=3)		Δ118-127	6.0 (0.0,n=3)
	M124A	6.0 (2.0,n=3		Δ128-137	9.5 (3.0,n=4)
	S125L	6.7 (2.3,n=3)		Δ138-147	8.0 (0.0,n=3)
	S125A	8.0 (0.0,n=5)		Δ148-157	9.3 (2.3,n=3)
	D126L	7.3 (1.2,n=3)		Δ118-132	9.0 (3.8,n=4)
	D126A	7.3 (1.2,n=3)		Δ123-137	8.0 (0.0,n=3)
	E127L	6.0 (2.0,n=3)		Δ128-142	10.0 (2.3,n=4)
	E127A	8.0 (2.8,n=5)		Δ133-147	9.3 (2.3,n=3)
Predicted α-helix	A138P	6.0 (2.3,n=4)		Δ138-152	5.0 (1.4,n=2)
	A144P	6.7 (1.2,n=3)		Δ143-157	4.8 (1.5,n=3)
	A148P	8.0 (2.8,n=4)		Δ118-137	7.3 (1.2,n=3)
	A154P	8.0 (0.0,n=3)		Δ138-157	5.5 (1.9,n=4)
Removal of negative charges	Δ126-128	8.0 (0.0,n=3)		Δ118-142	12.0 (0.0,n=3)
	Δ126-127	7.5 (1.0,n=4)		Δ123-147	7.6 (0.9,n=5)
	Δ127-128	8.0 (2.8,n=4)		Δ128-152	8.0 (0.0,n=3)
	Δ131-132	7.3 (1.2,n=3)		Δ133-157	6.0 (0.0,n=4)
	Δ137 Δ141	4.0 (0.0,n=3)		Δ118-147	10.0 (2.8,n=5)
	Δ131-132 Δ141	6.0 (2.0,n=3)		Δ123-152	13.6 (2.2,n=5)
	Δ149 Δ156	8.0 (0.0,n=3)		Δ128-157	7.3 (1.2,n=3)
	Δ126-128 Δ137 Δ141	8.0 (0.0,n=2)		Δ118-152	12.0 (0.0,n=3)
	Δ126-128 Δ131-132	8.0 (2.4,n=5)		Δ123-157	3 (0.0,n=3)
	Δ126-128 Δ131-132 Δ137 Δ141	10.4 (2.2,n=5)		Δ118-157	2.5 (0.6,n=4)
	Δ126-128 Δ131-132 Δ137 Δ141 Δ149 Δ156	10.0 (2.8,n=3)			
	Ins D129 E130 E131	6.7 (1.2,n=3)			
	D131A E132A	6.0 (2.3,n=6)			
	E137A E141A	6.7 (2.3,n=5)			
	D126A E127A E128A	7.2 (1.1,n=4)			
	D131K E132K	7.3 (1.2,n=3)			
	E137K E141K	6.5 (1.9,n=4)			
	D126K E127K E128K	9.0 (2.0,n=4)			
	D126K E127K E128K E137K E141K	8.0 (0.0,n=3)			
	D126K E127K E128K D131K E132K E137K E141K	10.0 (2.2,n=5)			

Next, the negatively charged portions of the loop region were investigated. Charged residues are known to be important for correct insertion and orientation of membrane proteins when translocated by the Sec machinery. This is particularly true for positive charges as explained by the positive inside rule; prevalence for lysine and arginine residues in the cytoplasmic loop regions (Heijne, 1986). This is not a counter balance of positive and negative charges as it is not the overall charge that has an effect, but the charge bias of lysine and arginine between loops in the cytoplasm and the periplasm. In fact, statistical studies have found that the frequency of negative charges is similar in translocated and untranslocated loop regions (Heijne, 1986, von Heijne & Gavel, 1988), indicating that they are less important for modulating activity of the Sec machinery. Nevertheless, it has been discovered that sufficiently large number of negative charges can have an effect on topology (Nilsson & von Heijne, 1990, Andersson *et al.*, 1992) and have been known to reduce the rate of translocation through the Sec machinery (Laws & Dalbey, 1989).

Therefore, it was hypothesised that manipulation of the conserved motif including some charged residues in the loop region may have an effect on the translocation of the polypeptide by the Sec machinery. Initially, residues in the M124, S125, D126 and E127 section (coloured in green in the sequence above Figure 4.4) were each individually mutated to alanine and to leucine. Figure 4.4 shows that none of these variants generated had any significant effect on Sec translocation of TMD3 as the M.I.C.s conferred by these constructs were, within experimental error, the same as that of 'wild-type' Rieske-Bla (M.I.C. of 8 µg/ml).

Closer inspection of the loop region revealed a number of negative charges dispersed throughout the sequence (boxed in orange in the sequence above Figure 4.4). Thus, a systematic approach was taken where each negatively charged residue was individually, or in multiples, deleted from the loop, or substituted to positively charged lysine residues. To this end, the following mutations were made Δ126-127, Δ127-128, Δ126-128, Δ131-

132, Δ137/Δ141, Δ149/Δ156, Δ131-132/Δ141, Δ126-128/Δ131-132, Δ126-128/Δ131-132/Δ137/Δ141, Δ126-128/Δ131-132/Δ137/Δ141/Δ149/Δ156, insertion of D129/E130/E131, D131A/E132A, E137A/E141A, D126A/E127A/E128A, D131K/E132K, D126K/E127K/E128K and D126K/E127K/E128K/E137K/E141K, D126K/E127K/E128K/D131K/E132K/E137K/E141K. Analysis of each of these variants showed that, with one exception, none of them resulted in a significant difference of the M.I.C. and consequently it can be inferred that there is no change to the Sec translocation of these fusion variants of Rieske-Bla. The sole exception was observed for the Δ137/Δ141-Rieske-Bla, which showed a significant decrease in M.I.C. to 4 µg/ml.

To determine whether the Δ137/Δ141-Rieske-Bla fusion protein was stably produced, this variant was expressed in the *tat*⁻ strain, membranes were prepared and the level of fusion protein assessed by Western blotting with an anti-Rieske antiserum. Figure 4.5 shows that there is potentially a slight decrease in protein produced that might account for the lower M.I.C. Moreover, if these same residues are deleted along with additional negatively charged residues (Δ126-128/Δ137/Δ141-Rieske-Bla) then the M.I.C. is no longer significantly different to Rieske-Bla. Taken together this indicates that these two negative charges at positions 137 and 141 are not critical for interaction with the Sec machinery and therefore the Δ137/Δ141-Rieske-Bla were not investigated further.

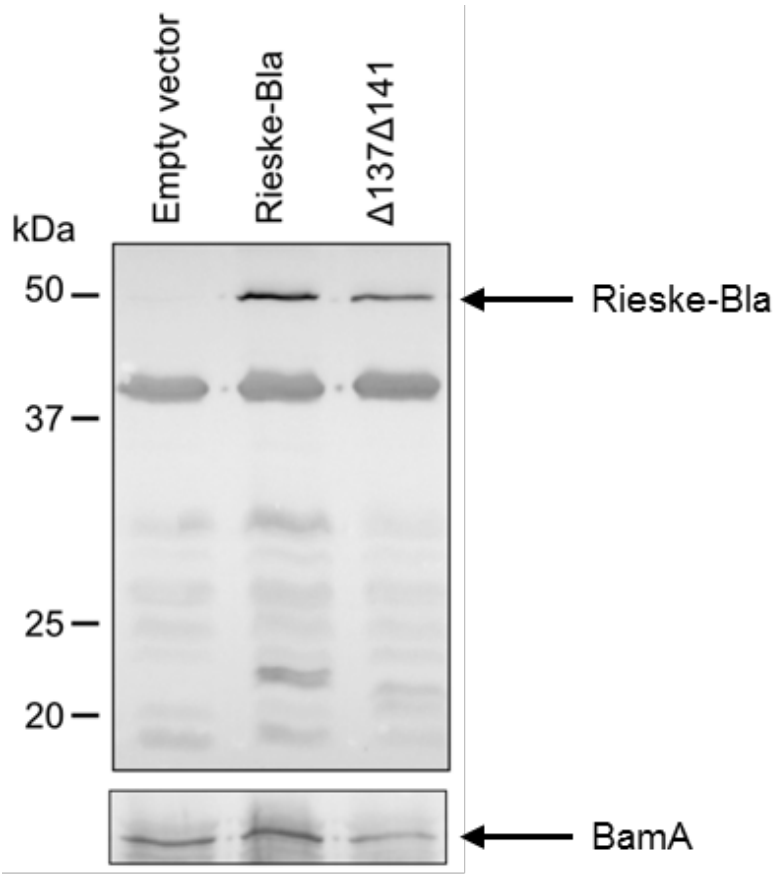


Figure 4.5 Stability of protein produced of variant Δ137/Δ141-Rieske-Bla.
Crude membrane extracts of DADE (*tat*⁻) harbouring the pSUPROM vector alone (Empty vector) or the pSUPROM vector encoding Rieske-Bla or loop truncation variant Δ137Δ141-Rieske-Bla was separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Rieske antibody. As a loading control the levels of the OM protein BamA were detected using an anti-BamA antiserum.

Finally, a potential role for the predicted α-helical region was probed. It should be noted that although there is not high sequence conservation across this region between Rieske proteins, the predicted helical structure in this region of the loop is conserved. To assess whether this potential helical region was essential for interaction with the Sec pathway, all of the alanines (A138, A144, A148 & A154, coloured pink in the sequence above Figure 4.4) were individually mutated to proline, an amino acid that is well known as a helix breaker (Nilsson & von Heijne, 1998, Nilsson *et al.*, 1998). As shown in Figure 4.4, none of these individual substitutions had any significant effect on the M.I.C. of ampicillin conferred by the fusion protein, indicating that interaction with the Sec pathway was not noticeably altered by these substitutions.

4.3.1.1.2 Truncations of residues in the cytoplasmic loop mostly have little effect on Sec translocation

Since the point mutagenesis failed to identify important determinants for Sec recognition of TMD3, more drastic mutations were constructed. To this end, truncations of the loop region were generated. Initially, sliding truncations of 5 amino acids at a time were constructed along the loop region; Δ118-122, Δ123-127, Δ128-132, Δ133-137, Δ138-142, Δ143-147, Δ148-152 and Δ153-157. As seen in Figure 4.4 and Table 4.1 most of these resulted in no significant change in M.I.C. compared to 'wild-type' Rieske-Bla, with the exception of the Δ128-132 truncation.

The Δ128-132 truncation resulted in a significant increase in M.I.C. compared to 'wild-type' Rieske-Bla (12 µg/ml) and is deleted for the following 5 residues; E128, V129, A130, D131 and E132. It is possible that the deletion of three negative charges could affect Sec translocation, although it has already been shown that removing all the negative charges from the loop has no significant effect on the M.I.C. of the variant (Δ126-128/Δ131-132/Δ137/Δ141/Δ149/Δ156-Rieske-Bla, 10 µg/ml) and neither did the removal of these specific negative charges (Δ126-128/Δ131-132-Rieske-Bla, with an M.I.C. of 8 µg/ml).

The next truncations to be generated were 10 residue deletions (Δ118-127, Δ128-137, Δ138-147 and Δ148-157). None of these variants showed any significant increase or decrease in M.I.C. (Figure 4.4), despite extending the deletion of the Δ128-132 region of Rieske-Bla that was previously shown to result in a significant increase in M.I.C. This indicates that this region is not critical for Sec recognition or release.

Subsequently, the sliding truncations were increased to 15 residues; Δ118-132, Δ123-137, Δ128-142, Δ133-147, Δ138-152 and Δ143-157. These variants also had no detectable effect on the Sec translocation of Rieske-Bla, with no significant differences observed in M.I.C conferred by these variants. Next, 20 residues were removed from the

loop region with two mutations, Δ118-137 and Δ138-157, again without any significant difference in M.I.C. compared to 'wild-type' Rieske-Bla.

When 25 residues were removed from the loop to generate truncations Δ118-142, Δ123-147, Δ128-152 and Δ133-157 there was no effect on M.I.C. conferred by the Rieske-Bla variants except for the Δ118-142 variant that yielded a small but significant increase in M.I.C. (12 µg/ml). Likewise, the removal of 30 residues from the cytoplasmic loop region (Δ118-147, Δ123-152 and Δ128-157) also identified a single variant, Δ123-152-Rieske-Bla, that supported a significantly increased M.I.C. of 13.6 µg/ml. The truncation of 35 residues from the cytoplasmic loop region, to give Δ123-157- and Δ118-152-Rieske-Bla showed that the Δ118-152-Rieske-Bla also supported an increase in M.I.C. to 12 µg/ml. This is in comparison to its counterpart Δ123-157-Rieske-Bla which supported a significantly decreased M.I.C. of 2 µg/ml.

Firstly, focusing on the variants that resulted in a statistically significant increase in M.I.C. (coloured green in the graph in Figure 4.4 and Table 4.1) it was reasoned that these variants may either be expressed at a higher level or have enhanced stability leading to more protein being present, or alternatively the truncations could remove feature(s) in the loop that partially blocks Sec translocation which would likewise result in a higher M.I.C. Although these variants showed increased M.I.C. overlap in their truncated regions, there is no obvious feature of the loop that they all delete which is not also truncated in other variants with 'wild-type' Rieske-Bla M.I.C. values. In the case of Δ123-152-Rieske-Bla and Δ118-152-Rieske-Bla it is possible that the increase in M.I.C. is a result of removing a large number of residues from the loop, although it should be noted that this effect is not seen with any of the other variants with 30 or 35 residues removed. The most drastic truncation made was a deletion of 40 residues (Δ118-157) leaving only 3 residues in the loop. As with the earlier deletion of 35 residues there was a substantial decrease in Sec translocation indicated by a significant change in M.I.C., Δ123-152-Rieske-Bla conferred an M.I.C. of 2 µg/ml and Δ118-157-Rieske-Bla an M.I.C. of 3 µg/ml.

These M.I.C. results are very similar to that seen with empty vector alone and would therefore be consistent with no Sec translocation of TMD3.

4.3.1.2 Rieske-Bla loop truncations affect Sec translocation

Next, an in-depth investigation into the truncation variants that specified a significant decrease in the M.I.C for ampicillin compared to 'wild-type' Rieske-Bla (Δ123-157-Rieske-Bla and Δ118-157-Rieske-Bla) was undertaken to try and understand why these variants conferred very little ampicillin resistance, particularly when the Δ118-152-Rieske-Bla resulted in an increased M.I.C. To this end, further variants were generated, where the truncation Δ118-152-Rieske-Bla was extended one residue at a time to give Δ118-153-, Δ118-154-, Δ118-155- and Δ118-156-Rieske-Bla.

Figure 4.6A is a representation of M.I.C.Evaluator™ test strip results obtained for the *tat*⁻ strain producing these truncation variants while Figure 4.6B is a graph of the mean M.I.C. with the s.d. as error bars. The Δ118-153-Rieske-Bla (depicted in green) showed a significant increase in M.I.C., whereas Δ118-154-Rieske-Bla showed no significant change of M.I.C. compared to Rieske-Bla (8 µg/ml; shown in purple). By comparison, Δ118-155-Rieske-Bla had a significantly reduced M.I.C. along with the variants Δ118-156-Rieske-Bla and previously seen Δ118-157-Rieske-Bla (in red).

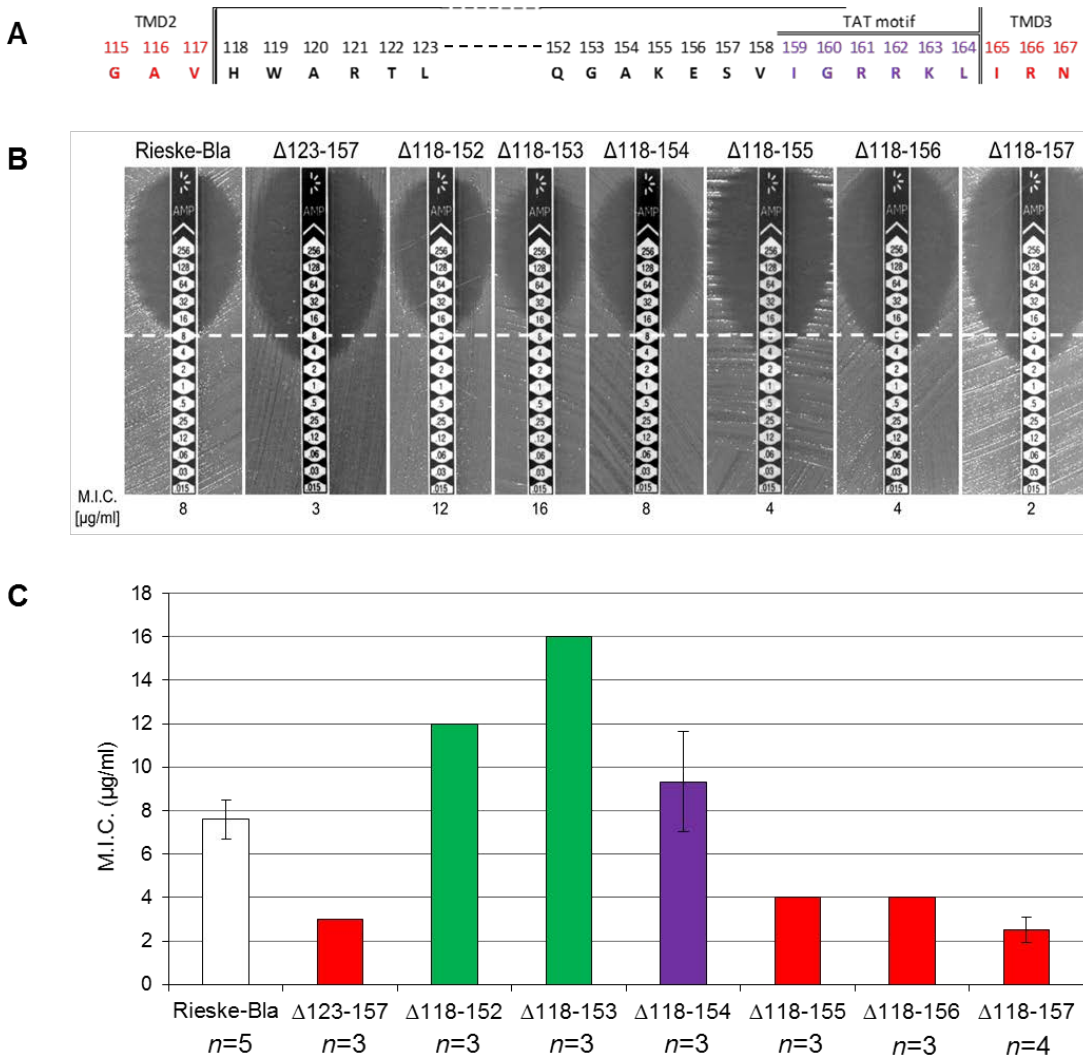


Figure 4.6 Significant M.I.C. differences mediated by closely related Rieske-Bla truncation variants.
A) The remainder of the conserved loop region for *S. coelicolor* Rieske for these truncation variants.
B) Representative M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding Rieske-Bla or loop truncation variants Δ123-157-Rieske-Bla, Δ118-152-Rieske-Bla, Δ118-153-Rieske-Bla, Δ118-154-Rieske-Bla, Δ118-155-Rieske-Bla, Δ118-156-Rieske-Bla and Δ118-157-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (μg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of Rieske-Bla (8 μg/ml).
C) The mean of the M.I.C.Evaluator™ assay data with the s.d. (error bars) was calculated and plotted on a graph. The number of biological replicates is indicated under the graph. Significance compared to Rieske-Bla is calculated by the Mann-Whitney Rank Sum Test and if $p<0.04$ then the bars are green (significant increase in M.I.C.) or red (significant decrease in M.I.C.), purple shows no significance calculated.

To confirm that the M.I.C. difference for the *tat*$^-$ strain producing the truncated fusion variants is due to an effect of the amino acid deletions and not a lack of protein production or protein stability, membranes from the *tat*$^-$ strain harbouring these variants were analysed by Western blotting with an anti-Rieske antiserum. Representative results are shown in Figure 4.7A, which gives an example of a Western blot of the truncated fusion proteins. It is apparent from the blot that a band shift is seen for the truncation variants due to shortening of the protein. For a comprehensive analysis, samples, collected in triplicate and analysed by Western blot, were quantified by densitometry analysis where Rieske-associated signals were normalised to the BamA-associated signals (which served as the loading control). The results were subsequently expressed as percentage of the normalised signal obtained for 'wild-type' Rieske-Bla, as seen in Figure 4.7B. The results indicate that all of the fusion proteins are stable and located in the membrane with a substantial amount of protein produced when analysed against Rieske-Bla. Therefore, it can be concluded that the severely reduced M.I.C.is due to effects of the truncations on interaction of TMD3 with the Sec pathway rather than on protein production or stability.

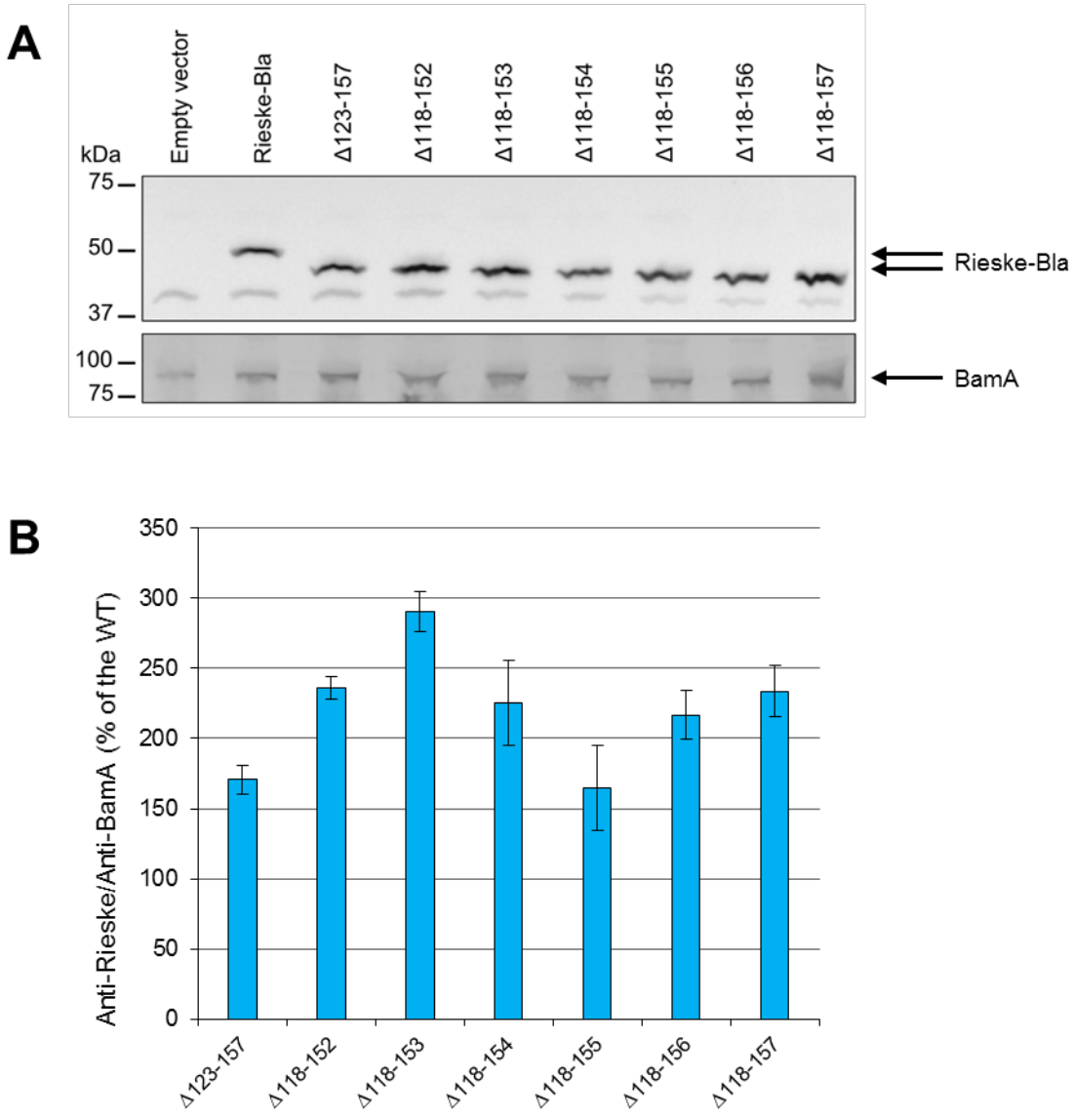


Figure 4.7 Protein is still stably produced from all truncation variants.
A) Crude membrane extracts of DADE (*tat*⁻) harbouring the pSUPROM vector alone (Empty vector) or the pSUPROM vector encoding Rieske-Bla or loop truncation variants Δ123-157-Rieske-Bla, Δ118-152-Rieske-Bla, Δ118-153-Rieske-Bla, Δ118-154-Rieske-Bla, Δ118-155-Rieske-Bla, Δ118-156-Rieske-Bla and Δ118-157-Rieske-Bla were separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Rieske peptide antibody. As a loading control the levels of the OM protein BamA were detected using an anti-BamA antiserum. Western blots used can be found in Figure 8.1.
B) Western blots were quantified by densitometry analysis and Rieske-associated signals were normalised with BamA-associated signals. The results were expressed as percentage of the normalised signal obtained for Rieske-Bla and the standard error of the mean (s.e.m) is indicated (error bars).

4.3.1.3 Addition of a positive charge restores Sec translocation to loop truncations

As seen above, a significant decrease in M.I.C. was found for the Δ118-155-Rieske-Bla construct, but intriguingly not for a one residue shorter truncation (Δ118-154-Rieske-Bla) that specified an M.I.C. similar to 'wild-type' Rieske-Bla. The additional residue that is deleted between these variants is K155. Previous data did not indicate that this residue *per se* was important for Sec translocation as other truncation variants containing a deletion of this residue had no significant effect on Sec translocation (for example, Δ153-157-Rieske-Bla with a mean M.I.C. of 9.6 μg/ml, Figure 4.4 and Table 4.1). However, an important Sec topology determinant is the positive inside rule; positively charged residues are four times more prevalent in the cytoplasmic loops of membrane proteins (Heijne, 1986). These residues probably remain in the cytoplasm due to the high energetic cost of translocating them across the membrane (Nilsson & von Heijne, 1990). Positive charges can influence the orientation of the TMD within membrane proteins and if charges within the loop regions are manipulated, the topology of some proteins, like EmrE, can be changed (Rapp *et al.*, 2006). Positive charges flanking TMDs can affect the insertion of TMDs, with similar effects from both arginine and lysine residues (Andersson *et al.*, 1992, Hessa *et al.*, 2007). This is particularly imperative when placed downstream of the helix where the charges can contribute to insertion (Lerch-Bader *et al.*, 2008).

The Δ118-155-Rieske-Bla construct no longer has any positive charges in the truncated cytoplasmic loop region, whereas Δ118-154-Rieske-Bla still contains one positive charge (K155). Assuming Rieske-Bla follows the Sec topology determinant, the positive inside rule, then removal of all the charges (particularly the positive charges) would be expected to have an effect on the insertion of its TMDs by the Sec machinery (Figure 4.8B). If this is the case then returning a positive charge to the truncated cytoplasmic loop region

within these variants should restore Sec translocation due to the return of a topological signal.

To this end, V158 was substituted with a lysine (V158K) and this mutation was introduced into ‘wild-type’ Rieske-Bla and the Δ118-153-, Δ118-154-, Δ118-155-, Δ118-156- and Δ118-157-Rieske-Bla truncation variants. As seen in Figure 4.8 the reintroduction of a positive charge into the ‘wild-type’ the Δ118-153- and the Δ118-154- Rieske-Bla, did not significantly alter the M.I.C. that they conferred. However, when it was introduced into the Δ118-155-, Δ118-156- and Δ118-157-Rieske-Bla constructs this ‘rescued’ the Sec translocation of TMD3 and restored the M.I.C. conferred by these variants to the same as ‘wild-type’ Rieske-Bla. Thus it can be concluded that the reduced M.I.C. from the original truncated variants was not due to the substantial decrease in size of the cytoplasmic loop but to the removal of a topological signal connected to the positive inside rule.

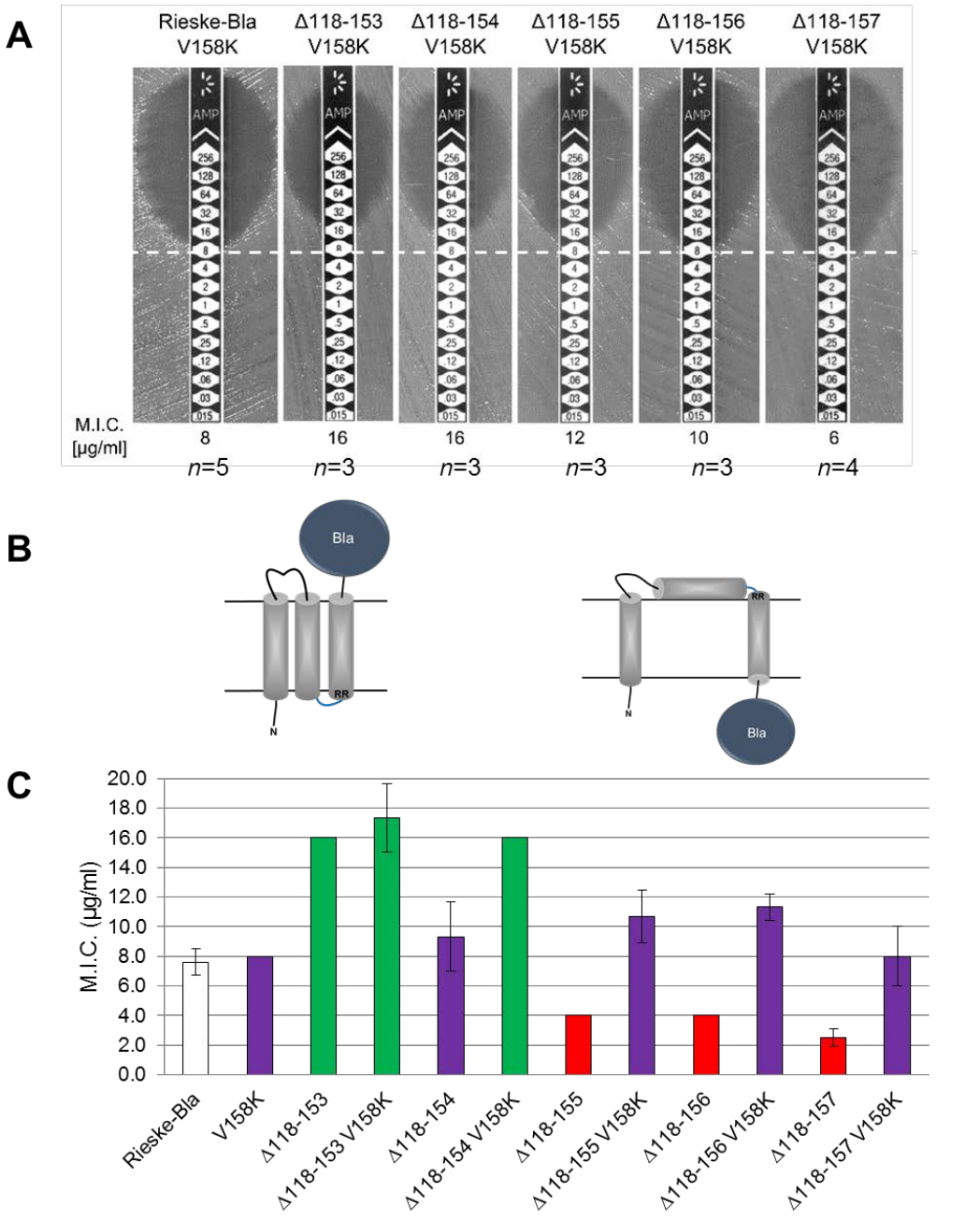


Figure 4.8 The addition of a positive charge to some of the loop truncations restores translocation of TMD3 through the Sec machinery.

A) Representative M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding Rieske-Bla or loop truncation variants with the mutation V158K, Δ118-153-V158K-Rieske-Bla, Δ118-154-V158K-Rieske-Bla, Δ118-155-V158K-Rieske-Bla, Δ118-156-V158K-Rieske-Bla and Δ118-157-V158K-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of Rieske-Bla (8 µg/ml). The number of biological replicates are indicated under the images.

B) Predicted topologies of Rieske-Bla with (left) and without (right) positively charged residues in the cytoplasmic loop region (blue).

C) The mean of the M.I.C.Evaluator™ assay data, with or without the mutation V158K, was plotted on a graph with the s.d. (error bars). Significance compared to Rieske-Bla is calculated by the Mann-Whitney Rank Sum Test and if $p<0.04$ then the bars are green (significant increase in M.I.C.) or red (significant decrease in M.I.C.), purple shows no significance calculated.

4.3.2 Topological determinants of Sec translocation

4.3.2.1 Positive charge distribution around TMD3 affects Sec translocation

During the systematic site-directed mutagenesis of the cytoplasmic loop region, a serendipitous mutation was introduced, R185C. This mutation was present within the construct Δ126-128-Rieske-Bla, and when this variant was tested for Sec-dependent translocation it was discovered that it had a strikingly high M.I.C. of 32 µg/ml. After re-sequencing the construct the additional mutation was found, consequently, both of these mutations were separately generated and re-tested for the M.I.C. of ampicillin to determine which one was the cause of the increase in Sec translocation. It was found that Δ126-128-Rieske-Bla produced a mean M.I.C. of 8 µg/ml, as indicated in Table 4.1, this is the same M.I.C. as 'wild-type' Rieske-Bla. On the other hand, R185C-Rieske-Bla generated a mean M.I.C. of 38 µg/ml (Figure 4.9), which is markedly high compared to Rieske-Bla (8 µg/ml) indicating that it is a potential Sec release over-ride variant.

The R185C mutation is located after the C-terminus of TMD3, and directly before the fusion junction with the reporter Bla. Although cysteines are generally tolerated at most positions in membrane proteins, it is not a substitution that is usually introduced during site-directed mutagenesis studies as it introduces an often-unwanted functionality. Therefore, R185 was also substituted by an alanine to observe whether the over-ride of Sec resulted from substitution of the arginine or by introduction of the cysteine. As seen in Figure 4.9 this R185A substitution further increased the M.I.C. to a mean of 64 µg/ml, indicating that it is the removal of the positively charged arginine residue that specifically causes increased efficiency of Sec-dependent export.

As discussed in Section 4.1.3, charged residues are important within membrane proteins to dictate the topology of the TMDs and are particularly prevalent in cytoplasmic loop regions, as described by the positive inside rule (Heijne, 1986). A change to the charges in loop regions, by adding or removing positively charged residues, can potentially reverse the topology of TMDs and sometimes the protein itself (Rapp *et al.*, 2006, Nilsson

& von Heijne, 1990). Charged residues are also important in the charge bias across a TMD and manipulation of this can also affect the orientation of a helix by increasing or decreasing the threshold hydrophobicity required for insertion (Ojemalm *et al.*, 2012).

'Frustrated' topologies occur when a sufficiently hydrophobic segment is prevented from inserting into the membrane by a neighbouring TMD with a strong orientational preference (Gafvelin & Vonheijne, 1994). This can also occur when a C-terminal positive charge prevents insertion of N_{out}-C_{in} TMD as the charge prevents the TMD from inserting 'incorrectly' by the positive inside rule. This frustrated TMD can be forced to insert if the hydrophobicity of the helix is increased (Ojemalm *et al.*, 2012).

The effect of R185 indicates that the Sec machinery samples the polypeptide before insertion which can permit these different topologies. This positive charge (R185) is located C-terminal to the helix and therefore, is at the periplasmic side of the membrane, where charges cause an energetic cost for translocation across the membrane against the transmembrane proton gradient (Yamagishi *et al.*, 2014). The loss of the positive charge changes the charge bias of TMD3, thereby, causing the Sec machinery to insert this variant more efficiently, yielding a higher M.I.C. for ampicillin.

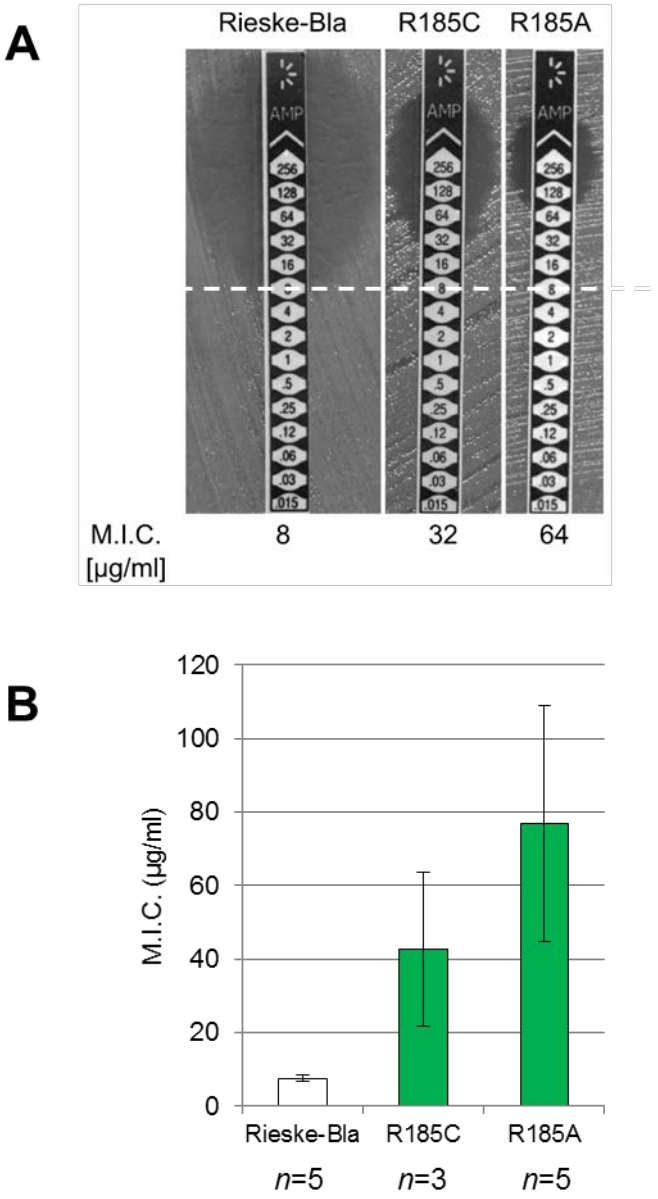


Figure 4.9 The removal of a positive charge C-terminal to TMD3 increases translocation through the Sec machinery.

A) Representative M.I.C.Evaluator™ strip test of DADE (*tat*⁻) harbouring the pSUPROM vector encoding Rieske-Bla or variants R185C-Rieske-Bla or R185A-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of Rieske-Bla (8 µg/ml).

(B) The mean of the M.I.C.Evaluator™ assay data with the s.d. (error bars) calculated and plotted on a graph. The number of biological replicates are indicated under the graph. Significance compared to Rieske-Bla is calculated by the Mann-Whitney Rank Sum Test, if $p<0.04$ then the bars are green (significant increase in M.I.C.).

4.3.2.2 Increasing the hydrophobicity of TMD3 increases Sec translocation

Another topological signal involved in the insertion and orientation of helices in the cytoplasmic membrane is the overall hydrophobicity of the TMD. Hydrophobicity is the driving force for insertion of a helix into the membrane, strongly hydrophobic segments entering the Sec translocon cause the lateral gate to open, which allows the helix to contact the lipid bilayer and escape into the membrane to generate a TMD (von Heijne, 1997, Hessa *et al.*, 2005, White & von Heijne, 2008). Quantification of TMD insertion into the mammalian ER (Hessa *et al.*, 2005, Hessa *et al.*, 2007), the yeast ER (Hessa *et al.*, 2009), the mitochondrial IM (Botelho *et al.*, 2011), a YidC-dependent protein in the *E. coli* IM (Xie *et al.*, 2007) and a TMD into the *E. coli* IM by SecYEG (Ojemalm *et al.*, 2013) have all been reported. These studies have allowed the generation of several position-specific biological hydrophobicity scales which measure the contribution of each amino acid on the free energy of insertion of a N_{out}-C_{in} TMD into the mammalian ER (Hessa *et al.*, 2007) and the bacterial IM (Ojemalm *et al.*, 2013). The reverse orientation (N_{in}-C_{out}) has also been investigated in the ER and the position specific effects of the individual amino acid residues are similar. However, the threshold hydrophobicity for a N_{in}-C_{out} TMD to be inserted is lower, this threshold is the level of hydrophobicity which is needed to open the lateral gate in the Sec translocon, thereby allowing N_{in}-C_{out} TMD to be less hydrophobic (Lundin *et al.*, 2008).

A highly hydrophobic helix has a strong orientational preference which can influence the topology of the nearby TMDs (Ojemalm *et al.*, 2012). Increasing the hydrophobicity of a TMD can change the orientation of a helix, previous findings have shown that an N-terminal signal-anchor sequence when increased in hydrophobicity will favour a N_{in}-C_{out} orientation (Goder & Spiess, 2003). The last TMD in a polytopic protein is on average more hydrophobic than central TMDs, comparable to the first TMD which is required to be strongly hydrophobic for targeting to the insertion machinery (Virkki *et al.*,

2014). This is supposedly due to the lack of neighbouring TMDs to exert an orientational preference.

To test how critical hydrophobicity is for Sec insertion of TMD3 of Rieske-Bla, an N_{in}-C_{out} TMD, mutations were generated to increase the hydrophobicity of TMD3 (P177L/S179L/G180L-Rieske-Bla, S179L/G180L-Rieske-Bla, S179L-Rieske-Bla and G180L-Rieske-Bla), these were then expressed in the *tat*$^-$ strain and tested for the M.I.C. of ampicillin conferred. Figure 4.10 demonstrates that increasing the hydrophobicity of TMD3 increases Sec translocation compared to Rieske-Bla, with up to a 25 fold larger M.I.C. This substantial increase in M.I.C. when only a few hydrophobic residues are added suggests that TMD3 is not particularly hydrophobic. The contribution to hydrophobicity in a TMD is usually additive so the more hydrophobic residues the greater the insertion propensity of the TMD (White & von Heijne, 2008).

The ΔG for TMD insertion into the ER membrane by Sec61can be calculated using a ΔG prediction server, http://dgpred.cbr.su.se/, based on the hydrophobicity scale from Hessa *et al.* (2007). This server can both calculate the ΔG_{app} for membrane insertion of a potential TMD and scan a protein sequence for putative TMDs, giving calculated ΔG_{app} for any TMDs found. In principle, if the ΔG value for a given potential helix is negative then it indicates that the sequence is recognised as a TMD by Sec61 and so would be integrated into the ER membrane. In comparison, if the value is positive it does not necessarily imply that the sequence is not a TMD, but it does flag that it is not a strongly hydrophobic helix and may not be efficiently inserted by itself suggesting that stabilising interactions from surrounding TMDs and residues are required. This would make it a marginally hydrophobic helix.

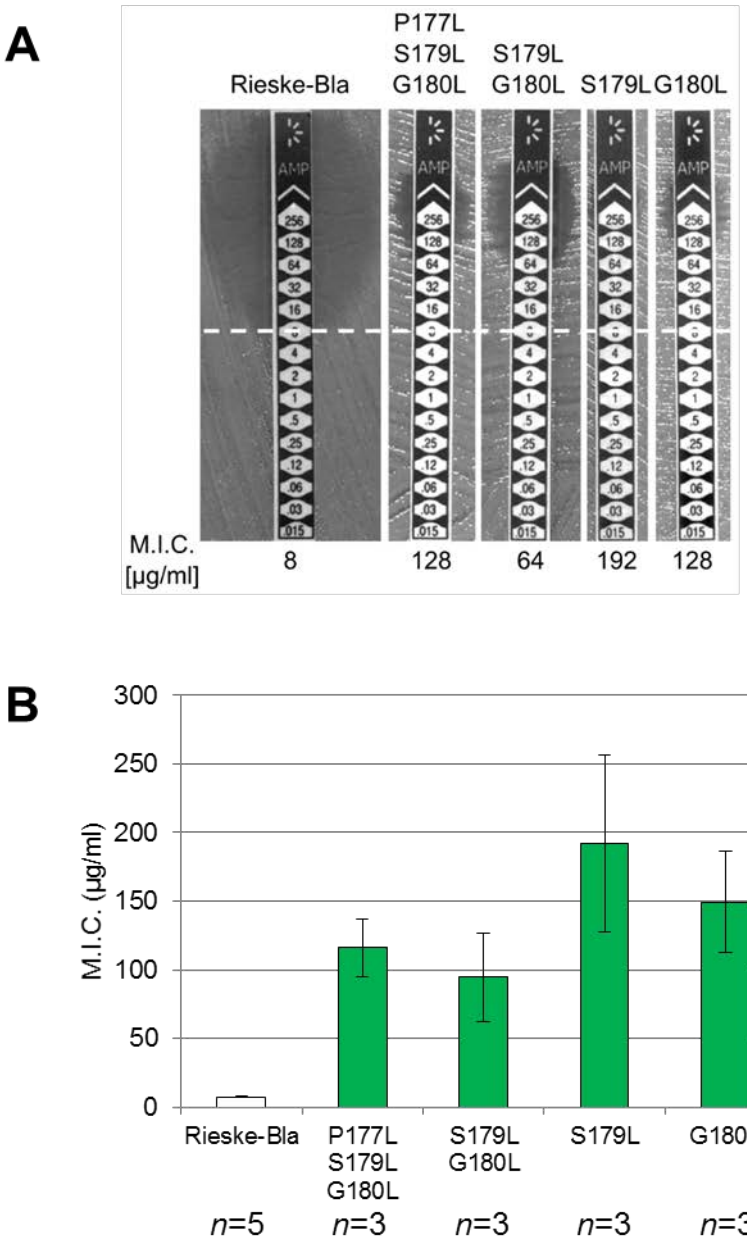


Figure 4.10 Increasing hydrophobicity of TMD3 increases translocation through the Sec machinery. A) Representative M.I.C.Evaluator™ strip test of DADE (*tat*⁻) harbouring the pSUPROM vector encoding Rieske-Bla or variants P177L/S179L/G180L-Rieske-Bla or S179L/G180L-Rieske-Bla.or G180L-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of Rieske-Bla (8 µg/ml).

(B) The mean of the M.I.C.Evaluator™ assay data with the s.d. (error bars) calculated and plotted on a graph. The number of biological replicates are indicated under the graph. Significance compared to Rieske-Bla is calculated by the Mann-Whitney Rank Sum Test if $p<0.04$ then the bars are green (significant increase in M.I.C.).

Using this server the ΔG_{app} for each TMD in *S. coelicolor* Rieske was calculated, and the results are given in Table 4.2. As these values are calculated for insertion into the ER membrane and Ojemalm *et al*. (2013) calculated that the threshold for the bacterial SecYEG complex is lower than the ER Sec61 complex, it is unlikely that TMD2, with the ΔG of 0.006, is a marginally hydrophobic helix. However, TMD3 does have a high positive value compared to TMD1 and TMD2, and this indicates that it might be a marginally hydrophobic helix. The ΔG_{app} of TMD3 of the S179L-Rieske-Bla variant is lower than that of 'wild-type' Rieske, which is in agreement with the M.I.C. results indicating that this variant is more readily inserted into the membrane.

Table 4.2 Predicted ΔG values for TMDs of Rieske.

	Sequence	Predicted ΔG
TMD1	VALLFTLSMLATIAFIAAFVAI	-2.177
TMD2	ISALNFALGMTLGVALFAIGAGAV	0.006
TMD3	IRNTMLGALTLVPLSGVVLLR	1.215
S179L-TMD3	IRNTMLGALTLVPL**L**GVVLLR	0.377

It is known that the last TMD of a polytopic membrane protein is usually strongly hydrophobic (Virkki *et al.*, 2014). As this is not the case for *S. coelicolor* Rieske, *E. coli* Sec-dependent IM proteins that have been experimentally shown to have three TMDs and N_{in}-C_{out} topology were analysed for the ΔG_{app} of their last TMD, shown in Table 4.3. These proteins were used because they have previously been analysed by Nilsson *et al.* (2000) who were investigating the reliability of topology prediction methods. The topology of these proteins has been agreed to be N_{in}-C_{out} by at least four out of five topology prediction methods, which increases the reliability of the orientation suggested. As seen in Table 4.3, the final TMD for all these proteins has a negative or close to zero ΔG value. This indicates that, unlike Rieske, TMD3 is fairly hydrophobic for all of these Sec-dependent proteins.

Table 4.3 Predicted ΔG values for *E. coli* IM proteins with three TMDs

Ecogene ID	Primary gene name	Predicted ΔG for TMD3
EG10332	***frdC***	-2.671
EG10934	***sdhD***	-3.173
EG10939	***secE***	-1.167
EG11695	***yidG***	-0.74
EG11696	***yidH***	-3.983
EG12395	***ybgE***	-1.234
EG12477	**yjeO**	-0.936
EG12575	***yjiN***	0.047
EG12747	***yhaH***	-3.097
EG12748	***yhaI***	-3.187
EG13028	***ygiZ***	-2.953
EG10178	***cyoA***	-2.775
EG10333	***frdC***	-1.43
EG10933	***sdhC***	-0.174
EG12382	***fxsA***	-1.359
EG12707	***ygbE***	-2.601
EG13987	***yniB***	-1.804

TMD3 of Rieske is shown to be a marginally hydrophobic TMD (ΔG value of 1.215, Table 4.2), which means it is not sufficiently hydrophobic to insert into the membrane on its own, requiring other TMDs or the influence of loop residues to increase the insertion potential. This low hydrophobicity could be significant in the release of the Rieske polypeptide by the Sec machinery, however, it cannot be the only feature for release as the Sec machinery is capable of inserting this helix to some degree.

4.3.3 Confirming the predicted topology of truncation Rieske-Bla fusion variants

Previously it was observed that the loss of topogenic signals prevents the insertion of TMD3 by the Sec machinery, highlighting the importance of the positive inside rule. Substituting positive charges can have the ability to change the orientation of helices inserted by Sec, switching loop regions from one side of the membrane to the other depending on where the charges are placed (Rapp *et al.*, 2006, von Heijne, 1989). The experiments described above show that truncation variant Δ118-155-Rieske-Bla is not inserted correctly as Bla is not located in the periplasm to confer resistance to ampicillin, but that the variant is located in the membrane so at least one of the TMD must be inserted. However, the overall topology of this construct is not clear. For example, is the loss of the topological signal K155 enough to re-orientate the helices and the loop regions, potentially generating a topologically frustrated TMD2 (Gafvelin & Vonheijne, 1994, Ojemalm *et al.*, 2012)? Or is TMD3 just left uninserted at the cytoplasm side of the membrane? Genetic experiments alone are not sufficient to answer these questions, and therefore, a sulphydryl labelling approach was undertaken to directly address which side of the membrane residues in the TMD2 and TMD3 loop regions are located for Rieske-Bla variants.

4.3.3.1 Experimental system for Sulphydryl labelling topology determination experiments

The accessibility of the cytoplasmic loop region was probed using cysteine labelling, according to the methodology described in Koch *et al.* (2012). The sulphydryl reactive agent used was methoxypolyethylene glycol maleimide (malPEG) which is a 5 kDa molecule that specifically reacts with an accessible sulfhydryl group to form an irreversible 3-thiosuccinimidyl ether linkage (Figure 4.11A). As a membrane-impermeable molecule it cannot cross the cytoplasmic membrane, therefore,

only reacts with cysteines exposed in the periplasm. Labelled cysteines are verified by Western blot and a covalent reaction with malPEG is detected by a 5 kDa band shift after immunoblotting.

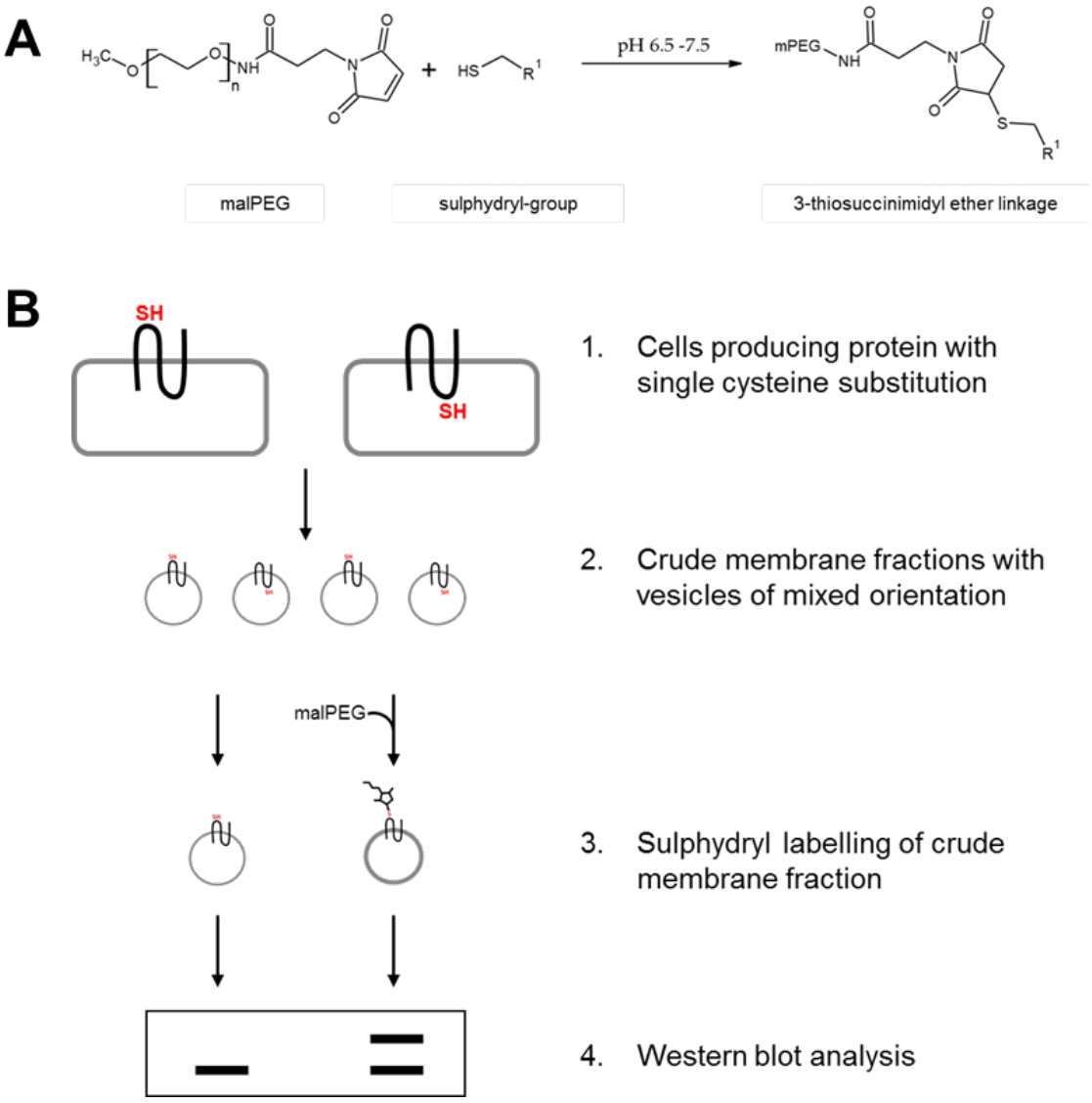


Figure 4.11 Reaction of malPEG and schematic to show approach to test for accessible cysteine residues.
A) The methoxypolyethylene glycol maleimide (malPEG) molecule reacts with an accessible sulfhydryl group to form an irreversible 3-thiosuccinimidyl ether linkage.
B) Schematic of sulphydryl labelling of crude membrane fractions. 1) Cells producing single cysteine substituted variants of Rieske-Bla were disrupted by ultracentrifugation isolating the crude membrane fraction. 2) The crude membrane fraction contains mixed orientation of membrane vesicles allowing both the periplasm and cytoplasm membrane protein domains to be exposed to the surrounding buffer. 3) Aliquots of the membrane fractions were either unlabelled or treated with malPEG. 4) Samples were separated by SDS-PAGE, electroblotted and immunoreactive bands were detected with an anti-Bla antibody. Labelling of malPEG was indicated by an electrophoretic mobility-shift compared to the band in the unlabelled sample. Labelled bands show the accessibility of cysteine residues to malPEG.

4.3.3.2 Selection of control residues for cysteine substitution to probe Rieske-Bla topology

'Wild-type' Rieske-Bla contains one cysteine residue which is buried within Bla and is therefore inaccessible to the surrounding buffer and labelling by malPEG. Cysteine substitutions were chosen within the loop regions of Rieske-Bla to detect labelling in the periplasm and the cytoplasm, as suitable controls for the experiment. V82 is in the periplasmic loop region between TMD1 and TMD2, consequently, this variant should be labelled when expressed in whole cells and treated with malPEG, Figure 4.12A. V158 is in the very C-terminal part of cytoplasmic loop region between TMD2 and TMD3 (seen in Figure 4.12A), therefore, in V158C-Rieske-Bla this cysteine should not be labelled by malPEG when expressed in whole cells. This residue is located outside of the truncations made in the cytoplasmic loop region and, therefore, it can be mutated in all the truncation variants to be tested.

It is critical that the cysteine substitutions generated are membrane-extrinsic and therefore potentially accessible for labelling. To test this, the protocol described in Figure 4.11B was followed. Crude membrane fractions of strains MC4100 (*tat*$^+$) or DADE (*tat*$^-$) producing either of these cysteine-substituted constructs, the 'wild-type' (i.e. Cys-less) Rieske-Bla or harbouring the empty plasmid vector were prepared by sonication and the membrane fraction isolated by ultracentrifugation. This produces membrane vesicles with mixed orientation exposing both periplasmic and cytoplasmic membrane protein domains to the surrounding buffer. The crude membrane fractions were split into two aliquots; one was left untreated (the negative control), whilst the other was treated with malPEG which would react with the accessible cysteine generating an additional mass to the truncation variant. After quenching the reaction with dithiothreitol (DTT), samples were separated by SDS-PAGE and analysed by Western blotting. As shown in Figure 4.12B following treatment with malPEG both of the cysteine variants were labelled,

indicated by a band shift. This confirms that both cysteines are accessible to the surrounding buffer.

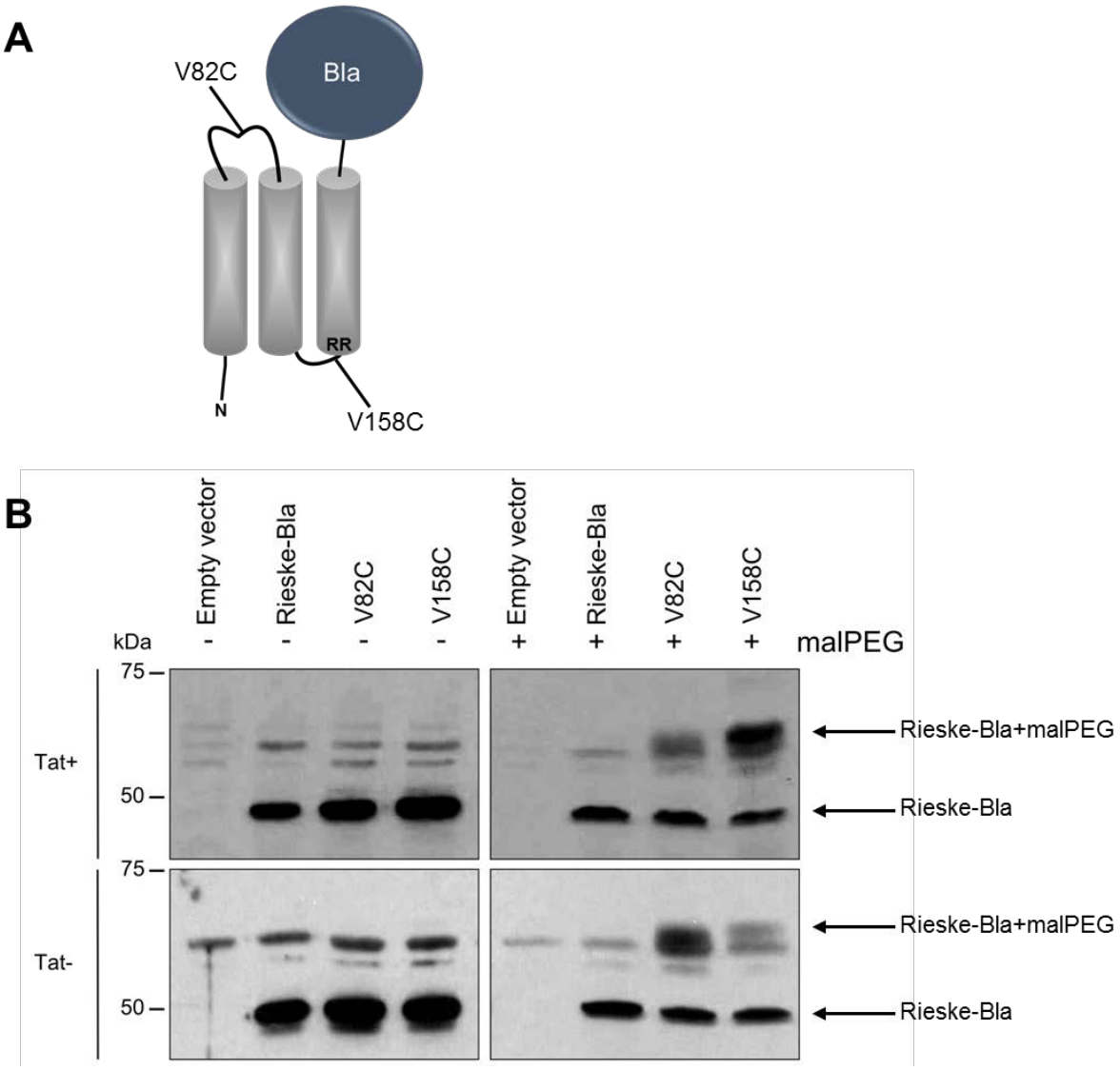


Figure 4.12 Analysis of cysteine accessibility by malPEG labelling in membrane fractions.
A) A schematic of where the control cysteine mutations are located on Rieske-Bla.
B) Labelling reactions were performed on membrane fractions of MC4100 (*tat*$^+$) or DADE (*tat*$^-$) strain harbouring an empty vector (pSUPROM) or the pSUPROM vector encoding Rieske-Bla or cysteine variants V82C-Rieske-Bla and V158C-Rieske-Bla. Samples were separated on SDS-PAGE, electroblotted and immunoreactive bands were detected with an anti-Bla antibody.

4.3.3.2.1 Determining the topology of 'wild-type' Rieske-Bla in whole cells

Next it was necessary to probe whether the introduced cysteines were located at the cytoplasmic or periplasmic side of the membrane, using sulphydryl labelling of whole cells. In this approach, one sample of whole cells is left untreated as a negative control whereas the other is incubated with malPEG to test the periplasmic accessibility of cysteine residues. Note that EDTA was present in both the positive and negative control samples as it is essential to render the cell envelope permeable to malPEG by chelating cations that normally stabilise the LPS component of the OM (Ferris & Beveridge, 1986, van Alphen *et al.*, 1978). This overall strategy is summarised in Figure 4.13A.

As these experiments are undertaken in whole cells, a cytoplasmic control is required to confirm that the cytoplasmic membrane was not disrupted during the procedure. The integrity control chosen was the HA-tagged SoxY protein (Figure 4.13B), that was derived from the single-cysteine-containing *Paracoccus pantotrophus* SoxYZ complex (Koch *et al*., 2012). The signal sequence of SoxY has been replaced with a HA-tag, trapping the complex in the cytoplasm. Therefore, if the cells are leaky during the experiment, HA-SoxY will be accessible to malPEG and a size shift will be observed for HA-SoxY in labelled samples.

Whole cell samples of strains MC4100 (*tat*$^+$) or DADE (*tat*$^-$) producing either of these cysteine-substituted constructs, the 'wild-type' (i.e. Cys-less) Rieske-Bla or harbouring the empty plasmid vector were treated with malPEG and the samples analysed (Figure 4.14). The results showed that the V82C-Rieske-Bla variant is labelled with malPEG in the *tat*$^+$ and *tat*$^-$ cells, indicating that this cysteine is located in the periplasm, whereas V158C remained unlabelled in both strains confirming a cytoplasmic location. These results support the expected topology of TMD1 and TMD2 of Rieske and confirm the findings of Keller *et al*. (2012) that there is no role for the Tat pathway in the membrane assembly of this portion of the protein.

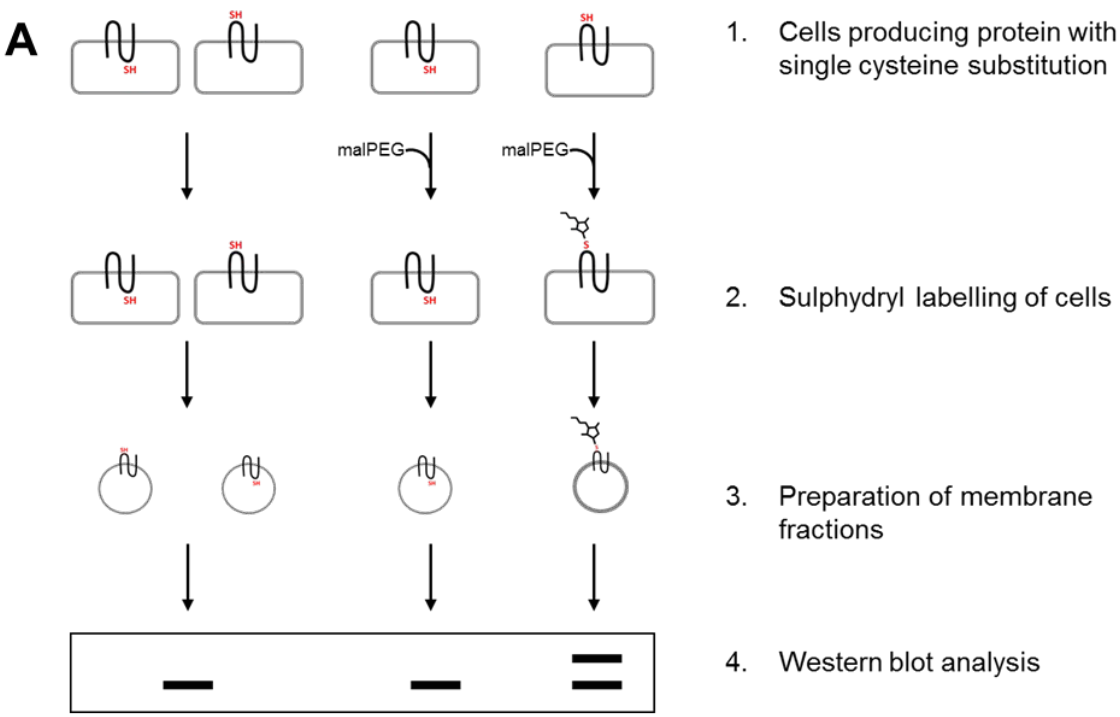


Figure 4.13 Schematic of sulphydryl labelling of whole cells with malPEG.
A) Sulphydryl labelling of periplasmic and cytoplasmic cysteines with malPEG 1) Aliquots of the cells producing single cysteine substituted variants of Rieske-Bla were treated in two ways. 2) The first aliquot was unlabelled, the second was labelled with malPEG. 3) The reaction was stopped and cells were disrupted by ultracentrifugation isolating the labelled or unlabelled crude membrane fraction. 4) Samples were separated by SDS-PAGE, electroblotted and immunoreactive bands were detected with an anti-Bla antibody. Labelling of malPEG was indicated by an electrophoretic mobility-shift compared to the band in the unlabelled sample. Labelled bands show the accessibility of cysteine residues to malPEG. B) The crystal structure of the SoxYZ complex of *Paracoccus pantotrophus* (Sauve *et al.*, 2007). The cytoplasmic control of cell integrity is SoxY which contains a cysteine residue at position 110. The mature SoxY protein has previously had a hemagglutinin tag fused to the N-terminus allowing detection by anti-HA immunoblotting (Koch *et al.*, 2012).

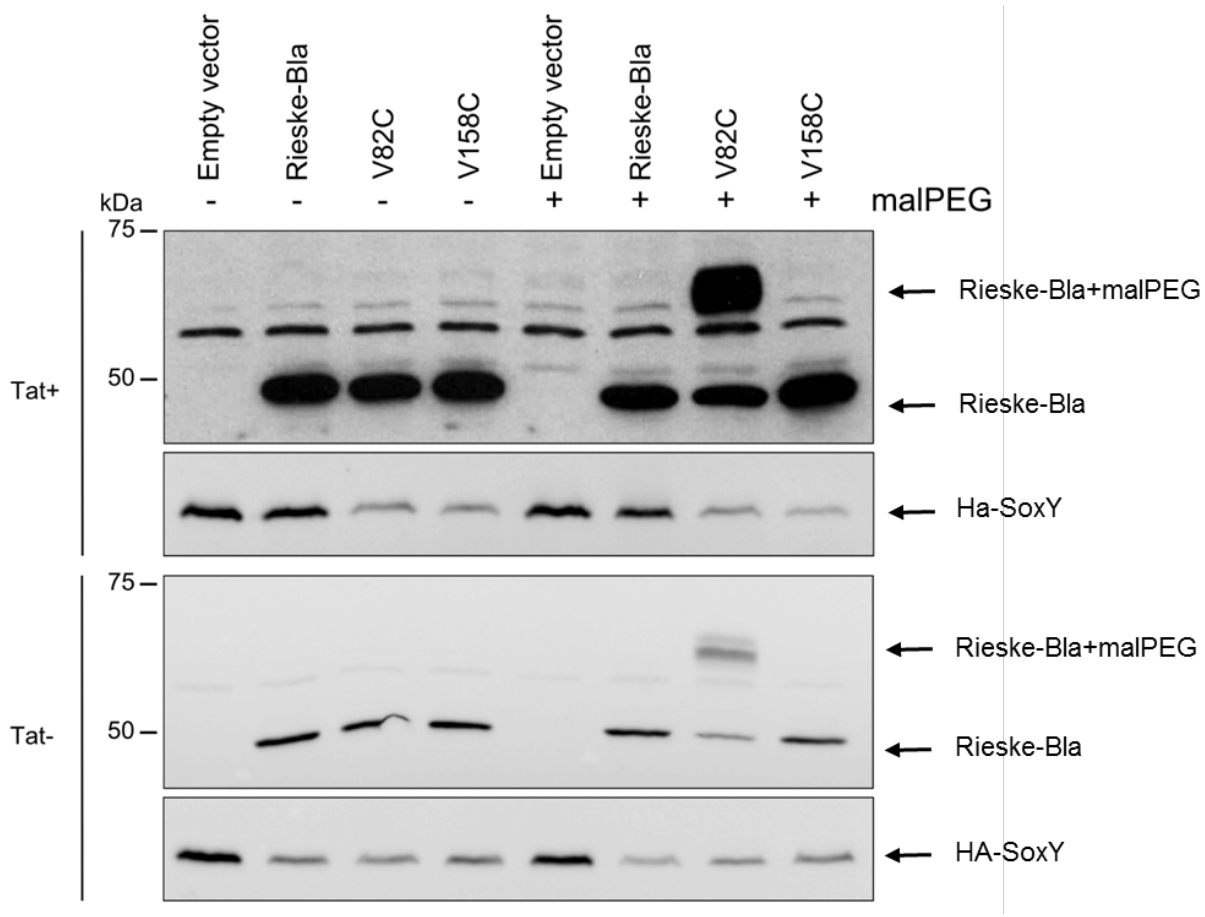


Figure 4.14 Analysis of cysteine accessibility by malPEG labelling in whole cells.
Labelling reactions were performed on whole cells of MC4100 (*tat*$^{+}$) or DADE (*tat*$^{-}$) strains harbouring an Empty vector (pSUPROM) or the pSUPROM vector encoding Rieske-Bla or cysteine variants V82C-Rieske-Bla and V158C-Rieske-Bla. Membrane fractions were isolated by ultracentrifugation and samples were separated on SDS-PAGE, electroblotted and immunoreactive bands of Rieske-Bla were detected with an anti-Bla antibody (top panels) and SoxY with an anti-HA antibody (bottom panels).

4.3.3.2.2 Large truncations of the TMD2-TMD3 loop block accessibility of V158C

To probe the arrangement of potentially topologically frustrated Rieske-Bla variants induced by loss of positive charges in the TMD2-TMD3 loop region, the V158C substitution was introduced into Δ118-153-, Δ118-155-, Δ118-156-, and Δ118-157-Rieske-Bla. Initially sulphydryl labelling experiments were conducted on crude membrane fractions of strains MC4100 (*tat*$^+$) or DADE (*tat*$^-$) producing each of these variants, the 'wild-type' Rieske-Bla or harbouring the empty plasmid vector to assess accessibility of the introduced cysteine residue. As Figure 4.15 shows, when membrane fractions containing each of the cysteine variants were treated with malPEG there was no labelling, as no band shift was detected. This indicates that the cysteines in these variants are not accessible to the surrounding buffer, possibly because they are partially pulled into the membrane by the very short loop region. Unfortunately, this finding precluded any further labelling analysis in whole cells. Without a positive control for the maleimide then these results could also be due to failed experimental conditions.

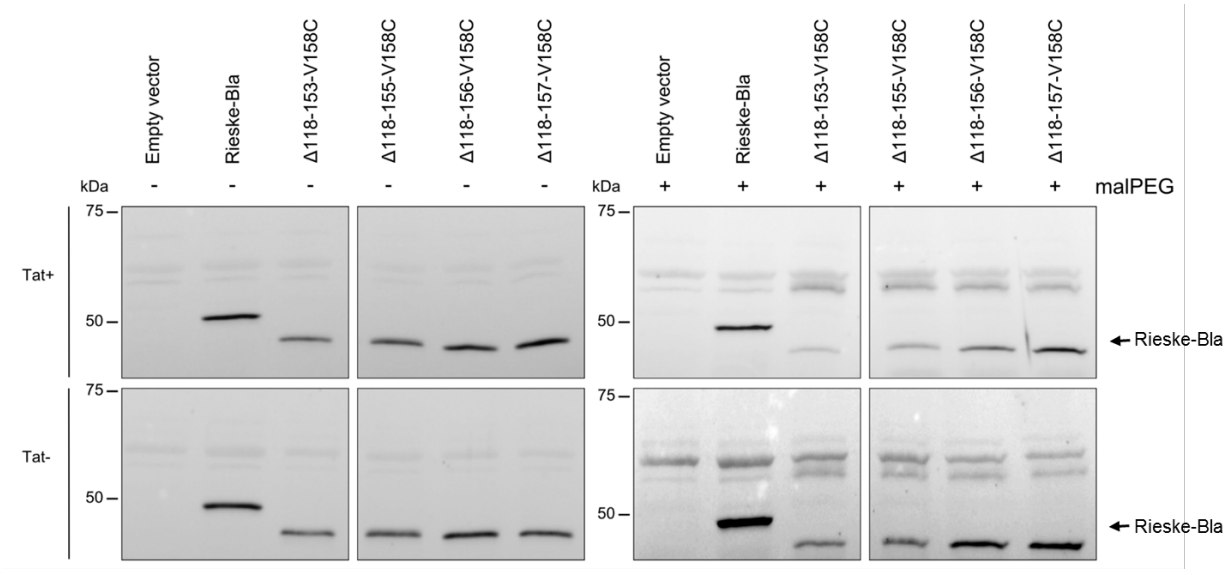


Figure 4.15 Western blot analysis of Rieske-Bla variants labelled with malPEG in membrane fractions.
Labelling reactions were performed on membrane fractions of DADE (*tat*$^-$) strain harbouring an empty vector (pSUPROM) or the pSUPROM vector encoding Rieske-Bla or cysteine variants. Samples were separated on SDS-PAGE, electroblotted and immunoreactive bands were detected with an anti-Bla antibody.

4.3.4 Positive charges are located C-terminal of TMD3 in actinobacterial Rieske proteins

As described previously, the positive inside rule is an important determinant for the correct insertion of TMDs by the Sec machinery. Positive charges have also been known to affect the topology of a protein even when acting over large distances. In the protein EmrE, a positive charge located in the C-terminus can invert the topology of the protein (Seppala *et al.*, 2010). Positive charges have also been found to anchor a marginally hydrophobic TMD even when placed 60 residues C-terminal of the helix (Fujita *et al.*, 2010). The importance of positive charges is also seen by topologically frustrated TMDs, where a positive charge forces the exclusion of a TMD, over-riding the hydrophobicity that drives insertion (Ojemalm *et al.*, 2012).

The results obtained so far in the first two results Chapters have shown that surprisingly there is some recognition of *S. coelicolor* Rieske TMD3 by the Sec pathway when it is fused to Bla and that increasing the hydrophobicity of TMD3 or removal of a positive charge shortly after the C-terminal end of TMD3 resulted in better insertion of TMD3 by the Sec machinery. With this knowledge in mind, a closer examination of actinobacterial Rieske proteins, particularly downstream of TMD3, was made, and an alignment of selected actinobacterial Rieske proteins covering TMD3 and immediately downstream is shown in Figure 4.16.

Interestingly, it is clear that for each Rieske protein examined there are a number of non-conserved positive charges located within the C-terminal vicinity of TMD3. It should be noted that these strings of positive charges are not found in other Rieske proteins that only contain a single TMD (Figure 3.1). Moreover, in the *S. coelicolor* Rieske-Bla construct used in this work these charges were removed as the Bla reporter is fused directly after R185 of Rieske. As the position and nature of the positive charges are not conserved among the actinobacterial Rieske proteins it is unlikely that they are essential for the structure/function of the Rieske protein. Instead it raised the possibility that they

are present to serve as a topological signal preventing the insertion of the helix by the Sec machinery.

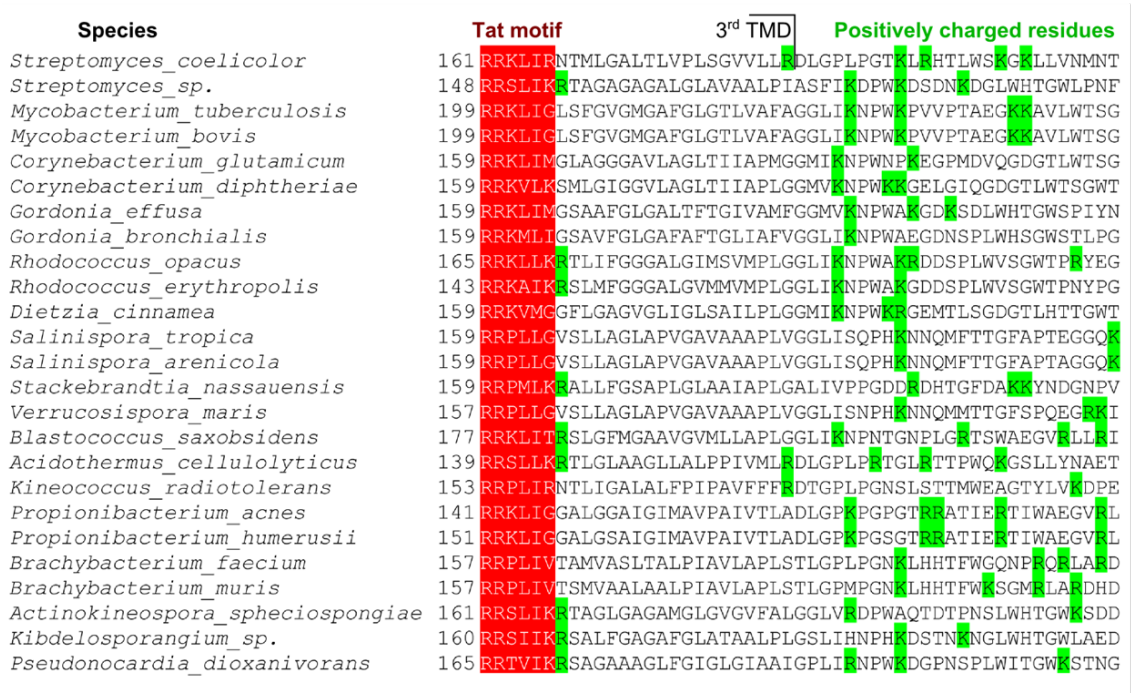


Figure 4.16 Alignment of TMD3 and downsteam regions of Actinobacterial Rieske proteins.
Protein sequence alignments of the TMD3 of the Rieske FeS proteins for different Actinobacteria. Proteins were aligned using Clustal WS and Jalview (Waterhouse *et al.*, 2009). The Tat motif is highlighted in red, positive charges are indicated in green and the predicted C-terminal end of TMD3 is also indicated.

To test this directly, a new construct was produced, termed 205-extended-Rieske-Bla, which included the first 205 residues of *S. coelicolor* Rieske (i.e. including 20 residues after TMD3 and all four positive charges, K194, R196, K202 and K204 shown in green in Figure 4.16). When this construct was expressed in DADE (*tat*$^-$) strain and tested for the M.I.C. of ampicillin, very unusual results were obtained. As shown in Figure 4.17A, the construct did not give reliable M.I.C. test results, instead sometimes there was a slight clearing of the bacteria down to an M.I.C. level of 1-2 µg/ml, with breakthrough colonies growing within the zone of clearing, suggesting that two populations of cells were present in the test. On other occasions, the cells grew up to the edge of the strip indicating they were resistant to the highest concentration of ampicillin present. These findings suggest that the cells producing this construct were stressed or unstable in some way.

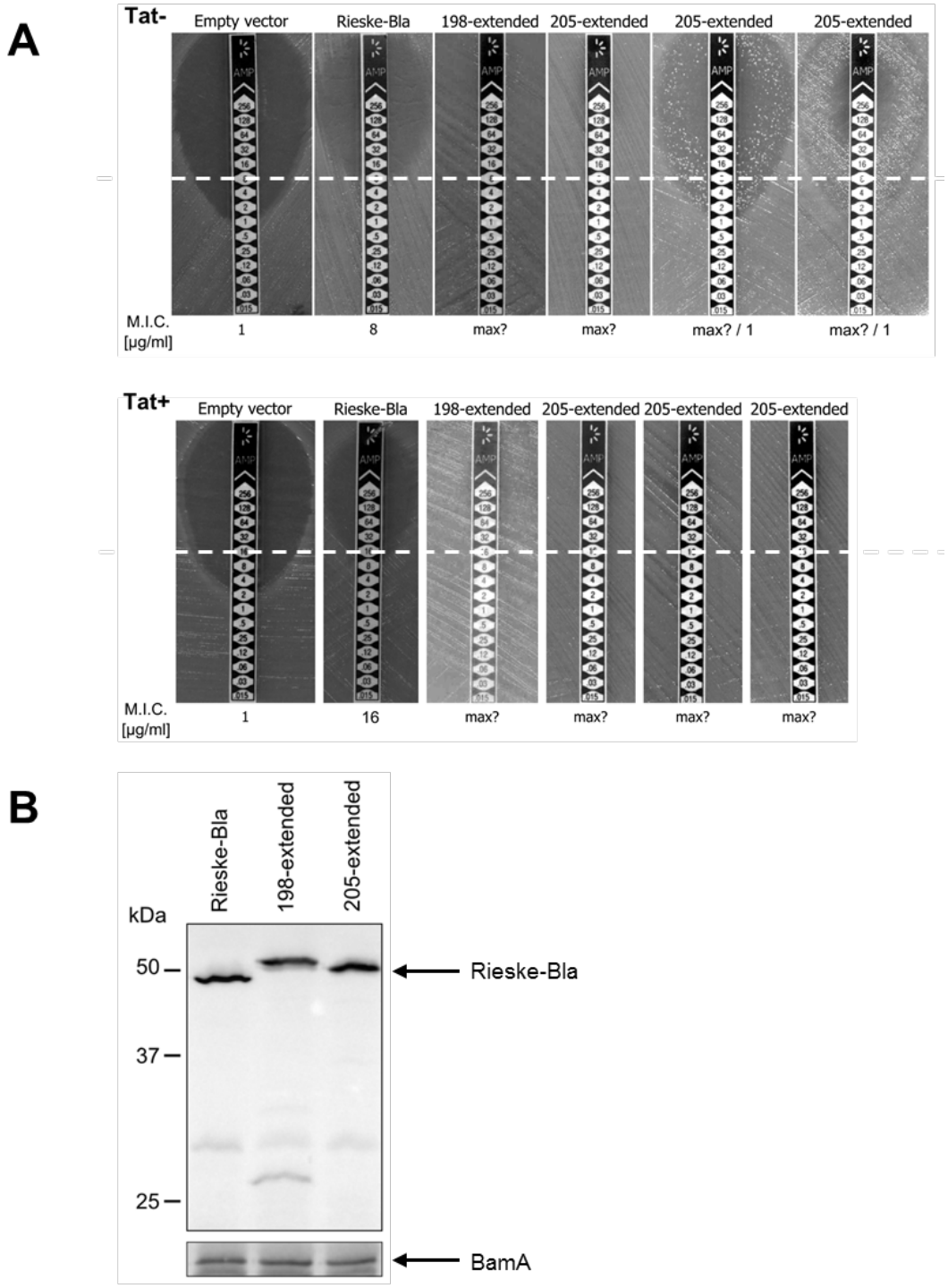


Figure 4.17 Extending the sequence of *S. coelicolor* Rieske-Bla leads to unexpected results.
A) Representative M.I.C.Evaluator™ strip test of MC4100 (*tat*$^+$) and DADE (*tat*$^-$) harbouring the pSUPROM vector (Empty vector) or the pSUPROM vector encoding Rieske-Bla or extension variants 198-extended-Rieske-Bla or three repeats of 205-extended-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of Rieske-Bla (*tat*$^-$ strain 8 µg/ml and *tat*$^+$ strain 16 µg/ml).
B) Crude membrane extracts of DADE (*tat*$^-$) harbouring the pSUPROM vector encoding Rieske-Bla or extended variants 198-Rieske-Bla and 205-Rieske-Bla, were separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with an anti-Bla antibody. As a loading control the levels of the OM protein BamA were detected using anti-BamA antiserum.

Given the unusual behaviour of cells harbouring the 205-extended-Rieske-Bla construct, a further construct was generated, 198-extended-Rieske-Bla, which contains an additional 13 residues compared to Rieske-Bla, including positive charges K194 and R196. Cells producing this construct did not show the unusual sectoring behaviour that was seen with the 205-extended-Rieske-Bla, but unexpectedly, it did confer an extremely high M.I.C. of > 256 µg/ml (Figure 4.17A). This level is very much higher than that conferred by positive charge removal or increased hydrophobicity of Rieske-Bla seen earlier in this Chapter and is completely contrary to what would be predicted.

The absence of the Tat machinery in the DADE strain produces a pleiotropic cell envelope phenotype which leads to a leaky cell envelope (Stanley *et al.*, 2001, Ize *et al.*, 2003). One possibility that was considered was that the extended Rieske-Bla protein could further interfere with the cell envelope somehow rendering cells resistant to ampicillin. To this end, these variants were also tested in MC4100 (the cognate *tat*$^+$ strain), however, Figure 4.17A shows that the M.I.C. values from these variants was also extremely high, with an M.I.C. of > 256 µg/ml. This indicates that the leaky cell envelope phenotype conferred by the loss of the Tat system is not the reason for this unexpected result.

To confirm that these atypical results are not due to overproduction of the Bla fusion proteins, or generation of truncated Bla-containing products, membranes from the *tat*$^-$ strain harbouring the extended-Rieske-Bla variants were analysed by Western blotting with an anti-Bla antiserum, the results are shown in Figure 4.17B. The levels of Rieske-Bla detected by the anti-Bla antiserum were not significantly different from 'wild-type' Rieske-Bla, and although there was a degradation band seen for 198-extended-Rieske-Bla, this is not seen in 205-extended-Rieske-Bla indicating that it is probably not the cause of the dramatic increase in M.I.C observed.

As the results obtained with the extended Rieske-Bla constructs were puzzling and difficult to explain, it was decided to test whether it was a unique feature of the *S. coelicolor* Rieske or whether a different actinobacterial Rieske protein would behave similarly. To this end, the *M. tuberculosis* Rieske sequence was codon optimised for expression in *E. coli* and the globular FeS domain was genetically removed and replaced with the mature region of Bla to generate *M. tuberculosis* Rieske-Bla. This 'wild-type' construct was expressed in strain DADE and tested for the M.I.C. of ampicillin it conferred. As seen in Figure 4.18, the M.I.C. was comparable to that conferred by *S. coelicolor* Rieske-Bla. To ensure that the *M. tuberculosis* Rieske-Bla was comparable in other aspects, a twin-arginine variant was generated with a mutation of R161A and R162D (RR-AD). The negative charge replacement of the twin-arginine in *M. tuberculosis* Rieske-Bla yielded a significantly decreased M.I.C. of 4 µg/ml, and is similar to previously shown data for *S. coelicolor* RRAD-Rieske-Bla (2 µg/ml; Figure 3.12).

The next construct generated, 243-extended-Rieske-Bla, includes the four positive charges located C-terminal of TMD3 (K228, K232, K241 and K242 seen in green in Figure 4.16). When this construct was expressed in the DADE (*tat*$^-$) strain and tested for the M.I.C. of ampicillin, it was found that the M.I.C. was decreased to 2 µg/ml. This indicates that, as expected, the positive charges beyond TMD3 act as a topological signal by preventing the Sec machinery from translocating Bla across the membrane.

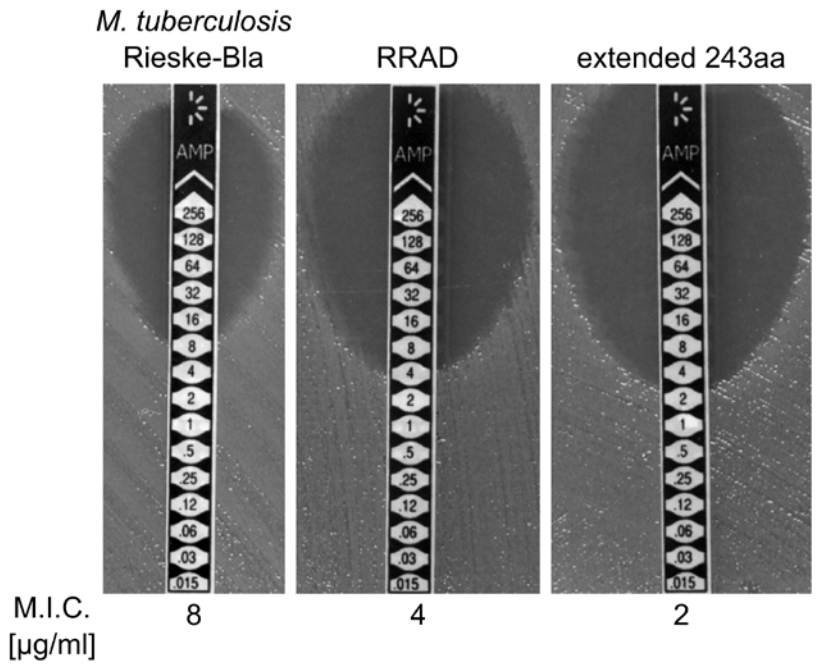


Figure 4.18 Extending the sequence of *M. tuberculosis* Rieske-Bla prevents translocation through the Sec machinery.
Representative M.I.C.Evaluator™ strip test DADE (*tat*) harbouring the pSUPROM vector encoding *M. tuberculosis* Rieske-Bla, the *M. tuberculosis* Tat motif variant RRAD-Rieske-Bla and the extension variant *M. tuberculosis* 243-extended-Rieske-Bla. Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The dotted white line across the pictures indicates the M.I.C. of *M. tuberculosis* Rieske-Bla (8 µg/ml).

4.4 Discussion

4.4.1 The cytoplasmic loop region is not important for the release of the Rieske polypeptide by the Sec machinery

The cytoplasmic loop region between TMD2 and TMD3 was predicted by Keller *et al.* (2012) to contain all the information required for the Sec machinery to release the Actinobacterial Rieske polypeptide. To test this prediction, mutations of conserved residues in this loop region were generated in the construct Rieske-Bla. It was found that even large truncations of the loop had no significant effect on Sec translocation of Rieske-Bla, allowing the conclusion that this cytoplasmic loop region does not contain any specific features that cause the Sec machinery to release the translocating polypeptide. It remains a possibility that the sequence of the cytoplasmic loop region is critically important for the recognition of TMD3 by the Tat machinery. The interaction of the Tat machinery with the cytoplasmic loop region will be investigated in Chapter 5.

4.4.2 The insertion of Rieske-Bla by the Sec machinery adheres to the topology determinant; the positive inside rule

The investigation of the cytoplasmic loop region revealed the dependence of the Rieske protein on the topological determinants necessary for correct insertion and orientation of proteins when translocated by the Sec machinery. The ability of a protein to be correctly inserted into the cytoplasmic membrane by the Sec machinery depends on a number of determinants within the TMDs; overall hydrophobicity, length and location of charged and polar residues (Hessa *et al.*, 2005, Hessa *et al.*, 2007). These are known as intrinsic determinants, there are also extrinsic determinants usually contained in the loop regions between TMDs, such as positive charges.

Charges are critical in orienting the TMDs of the protein. The positive inside rule proposes that positive charges are enriched in the cytoplasmic loops (Heijne, 1986) and

substitution of positive charges within the loop regions can change the topology of some proteins (Nilsson & von Heijne, 1990, Rapp *et al.*, 2006). In this Chapter it was seen that the substitution R185A, that removes a positive charge located at the C-terminal end of TMD3, dramatically increases integration of this helix by the Sec pathway. It has been previously reported that manipulation of the charge bias across a TMD can cause a change in the insertion of the TMD, even leading to 'frustrated' helices which are not inserted into the membrane (Ojemalm *et al.*, 2012). This charge could also enable a larger energetic cost for inserting this TMD (Yamagishi *et al.*, 2014). The removal of this positive charge at R185 has the effect of changing the charge bias and lowering the cost of insertion of this TMD, causing an increase in the M.I.C. specified by this variant (Figure 4.9).

The effect of the positive inside rule is also observed in the cytoplasmic loop region, Figure 4.6, where the removal of all of the positive charges in this loop (Δ118-155-Rieske-Bla) prevents the translocation of Bla across the membrane. Consequently, when a positive charge is returned into this region (Δ118-155-V158K-Rieske-Bla) the translocation of Bla into the periplasm is restored, Figure 4.8. This result indicates the dependence of Rieske-Bla on the positive inside rule as an important topological determinant required by the Sec machinery for the correct insertion and topology of helices in a Sec-dependent protein.

Manipulation of charges can cause incompatible insertion of TMDs, if they cannot sequentially follow the orientation of the previous helix due to the change in charge balance it can lead to topologically frustrated TMDs. These helices are then located outside of the membrane as their insertion is prevented by mislocalised positive charges (Ojemalm *et al.*, 2012). It is possible for the loss of positive charges in the cytoplasmic loop region of Rieske to generate a topologically frustrated TMD2, this would be difficult to determine as Bla would still be located in the cytoplasm and therefore, no M.I.C. for ampicillin would be seen. For this reason, labelling experiments were undertaken to

probe the subcellular location of the loop region between TMD2 and TMD3. However, a cysteine residue introduced into the truncated loop was not accessible to the malPEG labelling reagent, (Figure 4.15), probably due to the very short loop causing the cysteine to be positioned too close to the membrane. Thus the actual topology of these truncated variants remains to be experimentally determined.

4.4.3 TMD3 is a marginally hydrophobic helix

Hydrophobicity is the most important intrinsic determinant for insertion of a TMD by the Sec pathway. A helix is required to reach a threshold of hydrophobicity before the Sec machinery is able to release these residues into the membrane through the lateral gate. As discussed in Section 4.3.2.2, TMD3 of Actinobacterial Rieske proteins have low hydrophobicity, with a ΔG_{app} of 1.215 estimated for *S. coelicolor* Rieske TMD3, and so it is assumed that these are marginally hydrophobic helices. 20-30% of TMDs are marginally hydrophobic helices; they are not sufficiently hydrophobic to insert into the membrane by themselves (von Heijne, 2006). This is frequently due to a lack of hydrophobic residues but can also describe kinked, broken or short helices which are difficult for the Sec machinery to embed in the membrane. The insertion of these helices into the membrane is dependent on the adjoining TMDs and the residues outside of the helix (Hedin *et al.*, 2010, Lerch-Bader *et al.*, 2008). A neighbouring TMD with a strong orientational preference can either increase or decrease the insertion efficiency for the marginally hydrophobic helix (Ojemalm *et al.*, 2012). Adjacent helices are known to interact *via* polar and electrostatic interfaces (Buck *et al.*, 2007, Zhang *et al.*, 2007). It is presumably these interactions with TMD2 and the residues located in the cytoplasmic loop that allow the insertion of TMD3 of Rieske-Bla by the Sec machinery. Thus the substitution of a leucine residue into TMD3 significantly increases the hydrophobicity of the helix (ΔG_{app} of 0.377) allowing it to overcome the hydrophobicity threshold, which

allows a better efficiency of insertion into the membrane, observed by a higher M.I.C. for ampicillin.

It would be expected that substituting a couple of leucine residues would be additive to further increase the hydrophobicity of TMD3 and thus the M.I.C. conferred by the variant. This was not the case, as even when three leucine residues were substituted (P177L/S179L/G180L-Rieske-Bla) the M.I.C., although significantly increased from Rieske-Bla (8 µg/ml) was lower than the substitution of a single leucine residue (S179L-Rieske-Bla), Figure 4.10. It has been found that it is not only the number of hydrophobic residues that is important for the hydrophobicity but their location in the helix. Leucine residues located throughout the helix can be additive to the overall hydrophobicity, but clusters of leucines can diminish the propensity for insertion (Demirci *et al.*, 2013). This could explain why the M.I.C. for P177L/S179L/G180L-Rieske-Bla is lower than S179L-Rieske-Bla.

This data infers that the low hydrophobicity of TMD3 is a one of the factors that drives Sec release of the polypeptide during insertion. As TMD3 is the final helix it lacks the downstream TMD which would facilitate its insertion, and as seen in Table 4.3, the final TMD of a typical Sec-dependent protein is usually reasonably hydrophobic to ensure it is correctly inserted by the Sec machinery (Virkki *et al.*, 2014). Insertion efficiency of marginally hydrophobic helices can also be specifically adjusted by the manipulation of charged residues flanking the helix (Ojemalm *et al.*, 2012). In the case of Actinobacterial Rieske proteins a series of positive charges downstream of TMD3 appear to work in tandem with the low hydrophobicity of this helix to ensure that Sec does not insert this stretch of protein as a TMD, ultimately allowing it instead to be inserted by the Tat pathway.

4.4.4 Implications of this Sec-release mechanism for substrate recognition by the Tat pathway

Positive charges in membrane proteins are known to have long distance effects. A single positive charge located 60 residues C-terminal of a marginally hydrophobic helix can anchor it to the membrane with an N_{out}-C_{in} topology (Fujita *et al.*, 2010). This infers that the Sec machinery scans residues C-terminal of the translocating helix and responds to any signals that are present.

Ojemalm *et al.* (2012) discovered that if a TMD has a strong orientational preference due to positively charged flanking residues then the insertion of the downstream TMD can be affected. This TMD is possibly already inserted into the membrane or located in the lateral gate region of the Sec translocon when the strong orientational preference of the neighbouring TMD is determined by the Sec machinery. This causes the already inserted TMD to be forced across the membrane, with the final topology adopting a re-entrant loop, as seen in Figure 4.19A. Re-entrant regions are defined as a sequence that starts and ends on the same side of the membrane. They must penetrate the membrane to a depth between 1.5 and 25 Å and the more shallow insertions are required to have a 'clear turn' within the sequence (Viklund *et al.*, 2006). Analysis of these re-entrant regions within proteins has found that they are less hydrophobic than efficiently inserted TMDs (Yan & Luo, 2010).

It is, therefore, likely that the positive charges present C-terminal to TMD3 in Rieske (Figure 4.16) prevent its insertion as the charges will try to force an orientational preference on the helix. The Sec machinery should have already interacted with TMD3 possibly locating it in the lateral gate region when the positive charges are recognised, making it is very likely that TMD3 will form a re-entrant loop as the Sec machinery rejects the insertion, seen by the low M.I.C. in Figure 4.18. This is shown schematically in Figure 4.19B. As there is no further TMDs the Sec machinery presumably releases the Rieske protein at this point.

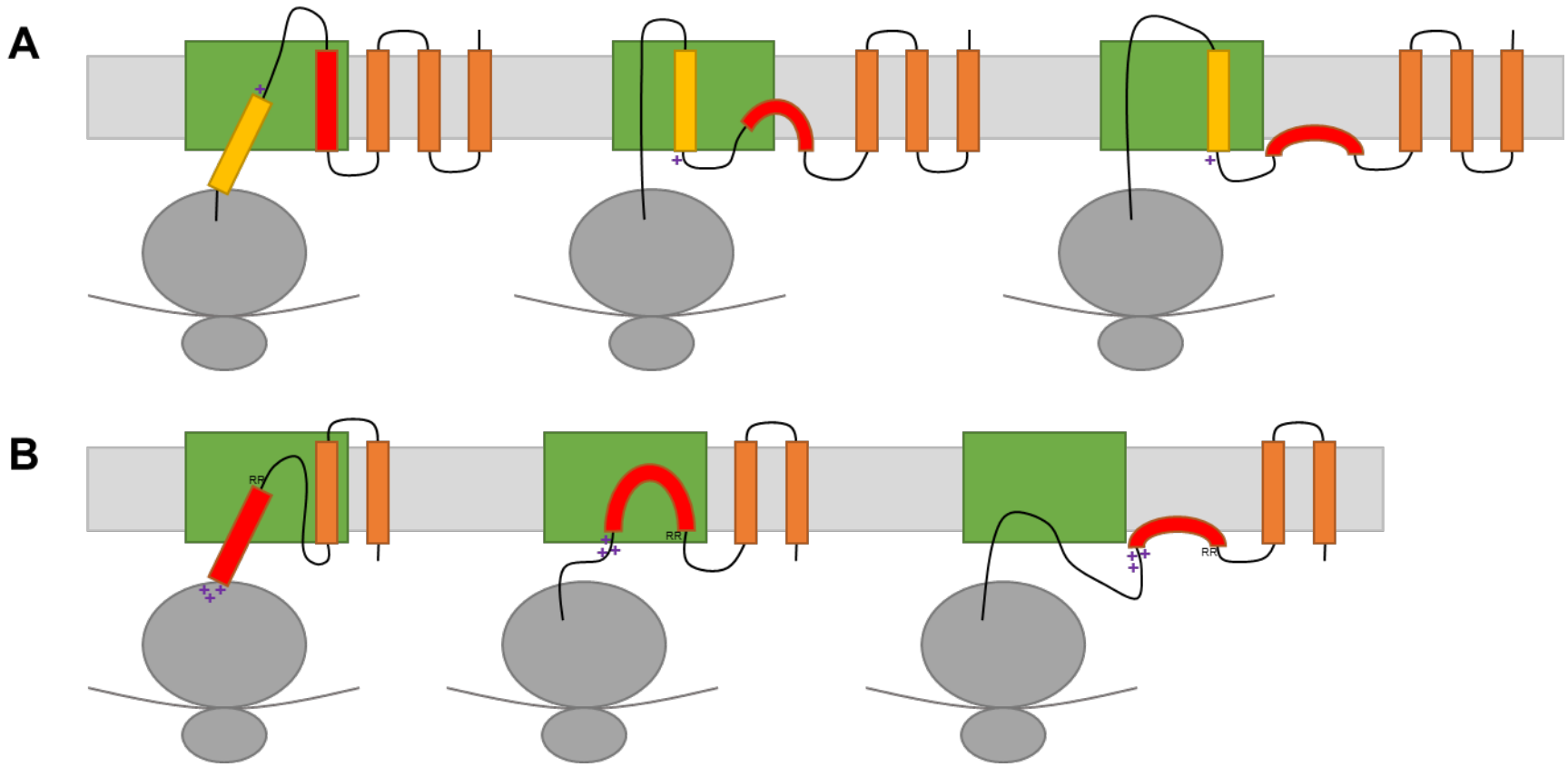


Figure 4.19 Schematics for the insertion of proteins by the Sec machinery.
A) Strong orientation preference of TMD can cause a re-entrant loop. The positive charge located C-terminal of the yellow TMD causes a strong orientational preference of N_{in}-C_{out} which forces the already inserted N_{in}-C_{out} neighbouring (red) TMD to reorient across the membrane forming a re-entrant loop on the cytoplasmic side of the membrane. Schematic based on Figure 6C (Ojemalm *et al.*, 2012).
B) Schematic of the insertion of the Rieske protein by the Sec machinery. TMD3 (red) is starting to be inserted by the Sec machinery with an N_{in}-C_{out} topology when the positive charges in the C-terminal are recognised and prevent the translocation across the membrane. This caused the TMD to be inserted across the membrane as a re-entrant loop and the Sec machinery releases the Rieske peptide.

Subsequently, the Tat machinery would recognise the twin-arginine motif and facilitate the insertion of TMD3. Clearly, this model would suggest that, in this case, the Tat machinery recognises a signal sequence that is partially located in the membrane, rather than free in the cytoplasm as has been generally assumed for most soluble Tat substrates. Interestingly, it might be that such a model can be applied more broadly to operation of the Tat pathway. It has long been known that Tat signal peptides often contain one or more positively charged residues in their c-regions which are known to act as a Sec-avoidance motif. Removal of these charges results in signal sequences that can now mediate transport by the Sec machinery (Bogsch *et al.*, 1997, Blaudeck *et al.*, 2003). In addition, signal peptides that direct proteins to the Tat machinery are known to be less hydrophobic than Sec signal peptides and if the hydrophobicity of a Tat signal peptide is increased it can mediate efficient transport by the Sec pathway (Cristobal *et al.*, 1999). Both of these features have been similarly noted above to promote release of Rieske TMD3 by the Sec machinery. However, despite possessing these 'Sec-avoidance' features, over half of the native *E. coli* Tat signal peptides are capable of transporting reporter proteins through the Sec pathway (Tullman-Ercek *et al.*, 2007). This raises the possibility that rather than being an exception, Sec interaction with Tat signal peptides is much more frequent, and that following abortive attempts at Sec-translocation, membrane-associated twin-arginine signal peptides are common substrates of the Tat pathway.

Several groups have proposed that signal peptide recognition by the Tat pathway always occurs in the lipid phase. For example, it has been shown that thylakoidal Tat substrates interact with the membrane before a subsequent interaction with Tat machinery (Musser & Theg, 2000, Asai *et al.*, 1999, Cline & Mori, 2001), and some Tat preproteins bind to the *E. coli* membrane in a *tat* mutant strain (Ray *et al.*, 2003). Using an *in vitro* model system, Tat precursors were found to have a tight interaction with lipid membranes, stabilised by both hydrophobicity and electrostatic interactions (Shanmugham *et al.*,

2006). Finally, direct interactions between the signal peptide and lipids has been found to easily interconvert to translocon binding which suggests that lipid-bound precursor proteins are functionally important in the Tat translocation pathway (Bageshwar *et al.*, 2009). However, in these cases it is not known whether the interaction with the membrane is partial or full penetration, or just an adsorption to the membrane surface.

Together the findings in this Chapter support the hypothesis that at least a fraction of Tat preproteins interact with the membrane as an early step in Tat translocation. It is possible that this binding is influenced by the Sec machinery trying to insert the signal peptide into the membrane before recognising the relatively low hydrophobicity and positively charged 'Sec-avoidance' motif and thus leaving the signal sequence as a re-entrant loop. This lipid-bound protein is then able to laterally diffuse along the membrane surface until it can bind to the Tat translocase for subsequent insertion into the membrane.

5 Tat translocation of *Streptomyces coelicolor* Rieske-AmiA

5.1 Introduction

5.1.1 Tat translocation and substrate recognition

The signal peptide is essential for targeting substrates to the Tat machinery and is usually located at the N-terminus of the protein (Berks, 1996). It comprises of a tripartite structure, seen in Figure 5.1, with a basic n-region, a hydrophobic middle h-region and a polar c-region. The n-region is of variable length and includes the essential consensus Tat motif, S**RR**xFLK. The twin-arginines in this sequence are crucial for recognition by the Tat machinery and are nearly always invariant (Berks, 1996). The h-region is slightly less hydrophobic than Sec signal peptides. The c-region contains a signal peptidase recognition site, usually AxA. In addition this region often contains a number of basic residues that together with the low hydrophobicity of the h-region acts as a Sec avoidance signal (Bogsch *et al.*, 1997, Blaudeck *et al.*, 2003, Cristobal *et al.*, 1999).

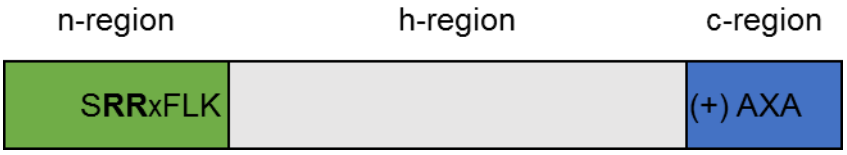


Figure 5.1 The tripartite structure of the Tat signal peptide.
The N-terminal Tat signal peptide has a tripartite structure; a basic n-region (green), a hydrophobic h-region (yellow) and a polar c-region (blue). The essential Tat motif is found within the n-region, whereas the signal peptidase cleavage site is located within the c-region.

Tat signal peptides are longer than Sec signal peptides which is largely due to an extended n-region which varies in length from a few residues up to 23 residues long, for example HyaA and DmsA (Tullman-Ercek *et al.*, 2007). This extended region is predicted to be α-helical and the primary sequence is generally conserved between related proteins from different bacterial species implying a functional importance.

There are a number of known signal peptide binding proteins that are believed to be important as proofreading chaperones or involved in quality control. For example, DmsD is required for cofacter insertion into apo-DmsA and TorD has been found to bind the

signal peptide of TorA, both of these interactions potentially prevent contact of the signal peptide with the Tat machinery (Sargent *et al.*, 2002, Oresnik *et al.*, 2001, Jack *et al.*, 2004). The n-region of the signal peptide is important in signal peptide-chaperone binding as the removal of this n-region abolishes TorD binding to the TorA signal peptide (Jack *et al.*, 2004).

Initially the signal peptide is recognised by and interacts with the TatBC complex (Cline & Mori, 2001). This selective binding is mediated by the twin-arginine motif of the signal peptide that contacts a surface patch on the cytoplasmic face of TatC, which comprises the N-terminal tail and loop region between TMD2 and TMD3 (Rollauer *et al.*, 2012). This has been experimentally confirmed by crosslinking (Zoufaly *et al.*, 2012) and mutational studies (Kreutzenbeck *et al.*, 2007, Strauch & Georgiou, 2007, Rollauer *et al.*, 2012) and enables the twin-arginines to interact with conserved glutamic acid residues of TatC (Berks *et al.*, 2014). It has also been found that the n-region can adopt a helical conformation which can associate with the intra-membrane face of TatC, the Lys of the signal peptide being proposed to interact with a conserved glutamate/glutamine located within the TatC cavity (Ramasamy *et al.*, 2013). If the n-region of the signal peptide is in an extended conformation, it is possible for it to interact with both the cytoplasmic side of TatC and the TatC cavity this also allows the important Phenylalanine of the signal peptide to stack with a conserved Phenylalanine of TatC (F94 in *E. coli* TatC) (Berks *et al.*, 2014).

Gerard *et al.* (2007) found two binding modes for signal peptide interacting with the thylakoid Tat machinery, peripheral-binding and deep-insertion. The peripheral-binding mode involved weak electrostatic interactions between the signal peptide and the TatBC complex allowing the signal peptide to still be accessible for proteolysis from the stroma. In comparison, when the signal peptide is in deep-insertion mode there is no proteolysis as the interactions are greatly strengthened with the signal peptide buried in the TatBC complex. The conversion from peripheral-binding to deep-insertion mode is driven by the PMF (Gerard & Cline, 2007).

This PMF-dependent conversion of binding has not been shown for the *E. coli* Tat machinery, however, weak-binding where the substrate can be exchanged and strong-binding where the substrate can be co-purified with the TatBC complex have been found for the same Tat substrate (Bageshwar *et al.*, 2009, Tarry *et al.*, 2009). The signal peptide has been found accommodated deep inside a TatC cavity in a hairpin loop conformation while the passenger domain is still located in the cytoplasm (Blummel *et al.*, 2015) Together this generates a model where the signal peptide is deeply bound within the Tat machinery and extends further into the cavity as the passenger domain crosses the membrane ultimately reaching the periplasm. Modelling has also suggested that the signal peptide can extend so it can both bind in the cytoplasm to TatC and expose its c-region signal peptidase site in the periplasm. This is supported by experimental observations which shows that TatC can act autonomously and guide the signal peptide across the membrane (Frobel *et al.*, 2012). The fate of the signal peptide after cleavage is unclear but it is probably released laterally into the membrane as with the uncleaved signal anchors of bacterial Rieske proteins.

5.1.2 The Rieske protein is Tat-dependent

Most well characterised Rieske proteins comprise of an N-terminal signal anchor followed by a globular FeS cofactor-containing domain. In bacteria and plants the N-terminal TMD comprises an uncleaved Tat signal sequence and it has been shown that the Tat pathway is essential for the integration of the Rieske protein into the membrane of bacteria and plant thylakoids (De Buck et al., 2007, Molik et al., 2001). Actinobacterial Rieske proteins contain 3TMDs with the final TMD encompassing a Tat signal sequence. The Tat machinery has been found to insert the final TMD of *S. coelicolor* Rieske in the heterologous host *E. coli* as well as in the native organism (Keller *et al.*, 2012, Hopkins *et al.*, 2014).

5.2 Aims

It was seen in Chapter 3, and previously shown by Keller *et al.* (2012) that the translocation of TMD3 of *S. ceolicolor* Rieske is Tat-dependent when AmiA is used as the fusion reporter. The aim of this chapter was to further assess the interactions of the Tat machinery with Rieske and dissect whether there are features within the cytoplasmic loop region which modulate the interactions with the Tat machinery, in particular the recognition of the Tat motif. To do this, a full programme of mutagenesis was undertaken with the Rieske-AmiA fusion, similar to that described in the previous chapter, and variants generated were tested for Tat-translocation by determining the level of growth on SDS-containing media. Variants of Rieske-AmiA that differed significantly from the 'wild-type' in terms of the amount of growth they supported on SDS-containing media were further analysed biochemically to support the results of the phenotypic observations.

5.3 Results

5.3.1 Investigating the cytoplasmic loop between TMD2 and TMD3

As discussed in Chapter 4, at the outset, the cytoplasmic loop region between TMD2 and TMD3 was predicted to contain all of the information necessary to co-ordinate the Sec and Tat machineries to facilitate the insertion of *S. coelicolor* Rieske (Keller *et al.*, 2012). In Chapter 4, however, it was seen that there are no clear features in this loop region that are important for the Sec machinery to release the Rieske polypeptide after the insertion of TMD1 and TMD2. However, it remains possible that the length and sequence features of this region are important for the recognition of the tethered signal sequence by the Tat machinery.

As indicated in Figure 4.2 there is some sequence conservation within the cytoplasmic loop region between Actinobacterial Rieske proteins. For example, the linker length is always 43 residues long residing between the C-terminus of TMD2 and finishing directly before the invariant twin-arginine of the Tat motif. There is a cluster of four negatively charged residues, a highly conserved RH motif and a predicted α-helical region. These features all appear to be conserved across the different Rieske proteins and, therefore, are expected to be important, perhaps with a potential role in presenting the signal sequence in the appropriate context to be recognised by Tat. In this Chapter the translocation of Rieske-AmiA by the Tat machinery was investigated.

5.3.1.1 Mutations generated in the cytoplasmic loop region had little effect on the insertion of TMD3 by the Tat machinery

The initial step in investigating whether the conserved residues in the cytoplasmic loop region were crucial for moderating interaction with the Tat pathway was to undertake a systematic site-directed mutagenesis approach where each of the conserved residues were individually mutated in the loop region encoded by the pSUPROM Rieske-AmiA construct. The cytoplasmic loop region is shown above Figure 5.2 and conserved

residues that were mutated in this Chapter are highlighted in colour. These variants were produced in MCDSSAC (*tat*$^+$) and MCDSSACΔtat (*tat*$^-$) strains to allow Tat-dependence to be scored, and tested for growth on SDS-containing media, as undertaken previously in Chapter 3, and a representative example is seen in Figure 5.2. The *tat*$^+$ and *tat*$^-$ strains expressing variants of the Rieske-AmiA fusion were grown until they reached stationary phase in liquid media, diluted to give approximately 10, 100, 1,000 or 10,000 cells per 5 µl aliquot and spotted onto standard LB media (as a control) and media containing 1% SDS. If the variant is Tat-compatible, and the Tat machinery is present (MCDSSAC strain) then there should be growth on SDS-containing media, and by comparison when there is no Tat machinery (MCDSSACΔtat strain) then there should be no growth. Table 5.1 shows the results seen for cytoplasmic loop variants when tested in the MCDSSAC (*tat*$^+$) strain, alongside the SDS growth for the twin-arginine variants that were previously described in Chapter 3 and the raw data can be found in the Appendix. Growth of variants on SDS-containing media is scored by a Y and no growth by an N.

5.3.1.2 Conserved amino acid substitutions had little effect on Tat translocation

The initial residues that were mutated in the Rieske-AmiA construct was the well conserved RH motif (R133 and H134, coloured green in the sequence above Figure 5.2). Aside from the twin-arginines of the Tat motif, these are the most conserved residues in the loop region. Two different substitutions were generated; R133H/H134R (RH-HR) and R133K/H134K (RH-KK). As shown in Figure 5.2 these substitutions did not change the growth on SDS-containing media compared to 'wild-type' Rieske-AmiA when expressed in the *tat*$^+$ strain, there was still growth when 10 cells per 5 µl was spotted. It can, therefore, be concluded that this motif is not essential for the Tat machinery to recognise the Rieske protein.

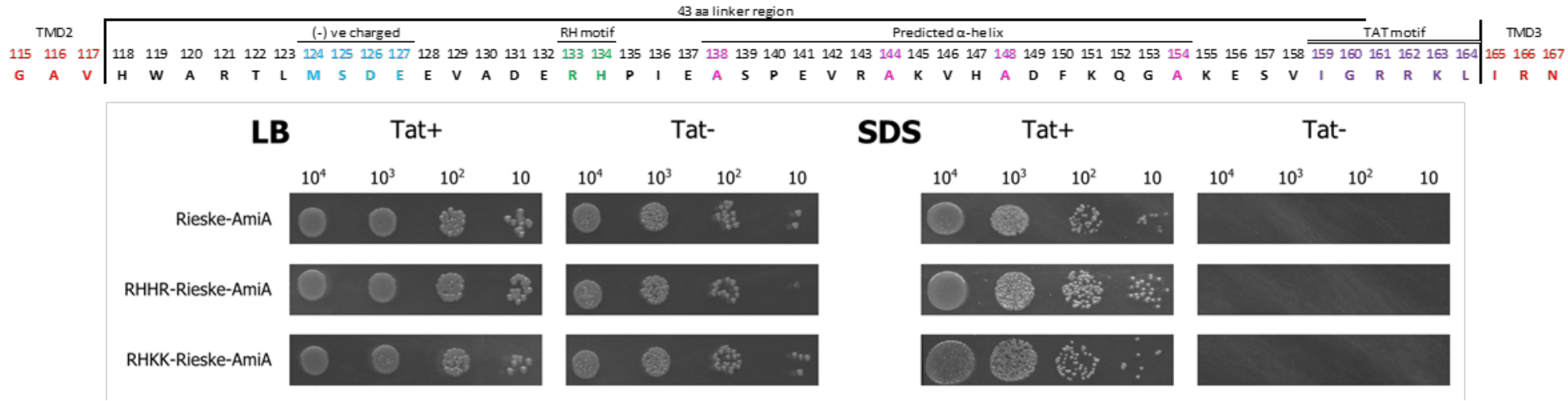


Figure 5.2 RHHR- and RHKK-Rieske-AmiA variants are Tat-dependent.
Representative spot tests of strains MCDSSAC (*tat*$^+$ strain) and MCDSSACΔtat (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), and the pSUPROM vector encoding Rieske-AmiA and RH motif fusion variants (RHHR-Rieske-AmiA and RHKK-Rieske-AmiA). Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto LB solid agar (left) and LB solid agar containing 1% (w/v) SDS (right). LB agar plates were incubated at 37°C for 16 h.
Above is the conserved loop region for *S. coelicolor* Rieske where the conserved residues that were mutated are indicated in colour.

Table 5.1 Tat dependence of Rieske-AmiA fusion variants.
Spot tests of strains MCDSSAC (*tat*$^+$ strain) and MCDSSACΔtat (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), or the pSUPROM vector encoding, Rieske-AmiA or the cytoplasmic loop variants. Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto LB containing 1% SDS. The *tat*$^+$ strain results were scored Y for growth and N (red) for no growth. Raw data can be found in the Appendix (Figure 8.3, Figure 8.4 and Figure 8.5).

	Variant	Growth on SDS		Variant	Growth on SDS
WT	Rieske-Bla	Y	Truncations	Δ118-122	Y
TAT motif	RRKK	Y		Δ123-127	Y
	RRAD	N		Δ128-132	Y
	RRAA	N		Δ133-137	Y
	RRKQ	N		Δ138-142	Y
	ΔRR	N		Δ143-147	Y
RH motif	RHHR	Y		Δ148-152	Y
	RHKK	Y		Δ153-157	Y
Negatively charged	M124L	Y		Δ158-162	N
	M124A	Y		Δ118-127	Y
	S125L	Y		Δ128-137	Y
	S125A	Y		Δ138-147	Y
	D126L	Y		Δ148-157	Y
	D126A	Y		Δ118-132	Y
	E127L	Y		Δ123-137	Y
	E127A	Y		Δ128-142	Y
Predicted α-helix	A138P	Y		Δ133-147	Y
	A144P	Y		Δ138-152	Y
	A148P	Y		Δ143-157	Y
	A154P	Y		Δ118-137	Y
				Δ138-157	Y
				Δ118-142	Y
				Δ123-147	Y
				Δ128-152	Y
				Δ133-157	Y
				Δ118-147	Y
				Δ123-152	Y
				Δ128-157	Y
				Δ118-152	Y
				Δ123-157	N
				Δ118-157	N

The subsequent region to mutate was the conserved negatively charged region. This contains four residues; M124, S125, D126 and E127 which are coloured in blue in the sequence above Figure 5.2. These were individually mutated to alanine and to leucine and tested for Tat dependence on SDS-containing media. When variants were expressed in the *tat*$^+$ strain there was no growth difference observed compared to 'wild-type' Rieske-AmiA, as indicated in Table 5.1.

The final conserved region is the predicted α-helix section. To assess if this structure is essential for Tat machinery recognition, the alanine residues in the loop (A138, A144, A148 and A154, coloured pink in the sequence above Figure 5.2) were mutated to prolines. As discussed previously, these mutations should disrupt any α-helix structure as proline is known to be a helix breaker (Nilsson & von Heijne, 1998, Nilsson *et al.*, 1998). Table 5.1 shows that there is no change in growth on SDS-containing media for these variants when expressed in the *tat*$^+$ strain, indicating that there was no change in the interactions with the Tat machinery.

5.3.1.3 Truncations of residues in the cytoplasmic loop had little effect on Tat translocation

As the site-directed mutagenesis of the cytoplasmic loop region had no detectable effect on the translocation of Rieske-AmiA by the Tat machinery, more extreme mutations, i.e. truncations of the loop region were generated. At the outset, sliding truncations of 5 aa at a time were constructed along the loop region to give Δ118-122-, Δ123-127-, Δ128-132-, Δ133-137-, Δ138-142-, Δ143-147-, Δ148-152-, Δ153-157- and Δ158-162-Rieske-AmiA. As seen in Table 5.1, and presented in the Appendix of this thesis, most of these variants supported no significant change in growth on SDS-containing media compared to 'wild-type' Rieske-AmiA when variants were expressed in the *tat*$^+$ strain. The exception is Δ158-162-Rieske-AmiA which did not support any growth on SDS-containing media. This result is exactly as predicted since these deleted residues (V158, I159, G160, R161 and R162) include the essential twin-arginine of the Tat motif. If these

residues are removed then it is not possible for the Tat machinery to recognise the tethered sequence and so there is no translocation of TMD3 or AmiA into the periplasm and no growth on SDS-containing media.

The subsequent truncations to be generated were deletions of 10 residues (Δ118-127-, Δ128-137-, Δ138-147- and Δ148-157-Rieske-AmiA). None of these truncated variants when expressed in the *tat*$^{+}$ strain and tested for growth on SDS-containing media exhibited a growth difference relative to 'wild-type' Rieske-AmiA. Therefore, the truncation was increased to 15 residues to give Δ118-132-, Δ123-137-, Δ128-142-, Δ133-147-, Δ138-152- and Δ143-157-Rieske-AmiA. Again, the expression of these variants in the *tat*$^{+}$ strain followed by testing on SDS-containing media also yielded 'wild-type' growth. Next, 20 and 25 residue truncations were tested by generating constructs Δ118-137-, Δ138-157-, Δ118-142-, Δ123-147-, Δ128-152- and Δ133-157-Rieske-AmiA. These also gave 'wild-type' Rieske-AmiA levels of growth when tested on SDS-containing media. Subsequently, 30 residues were removed from the cytoplasmic loop region (Δ118-147-, Δ123-152- and Δ128-157-Rieske-AmiA) and these variants also demonstrated Tat-dependent growth on SDS-containing media.

The first significant result was obtained when 35 residue truncations were constructed, generating Δ123-157- and Δ118-152-Rieske-AmiA. Although Δ118-152-Rieske-AmiA supported 'wild-type' growth, the Δ123-157-Rieske-AmiA did not allow any grow on SDS-containing media (Figure 5.3, Table 5.1). Therefore, a further deletion was constructed to remove 40 residues from the loop region, Δ118-157-Rieske-AmiA – this leaves only 3 residues remaining in the loop region (V158, I159 and G160 located immediately before the twin-arginines). Again, this variant also supported no growth on SDS-containing media (Figure 5.3, Table 5.1).

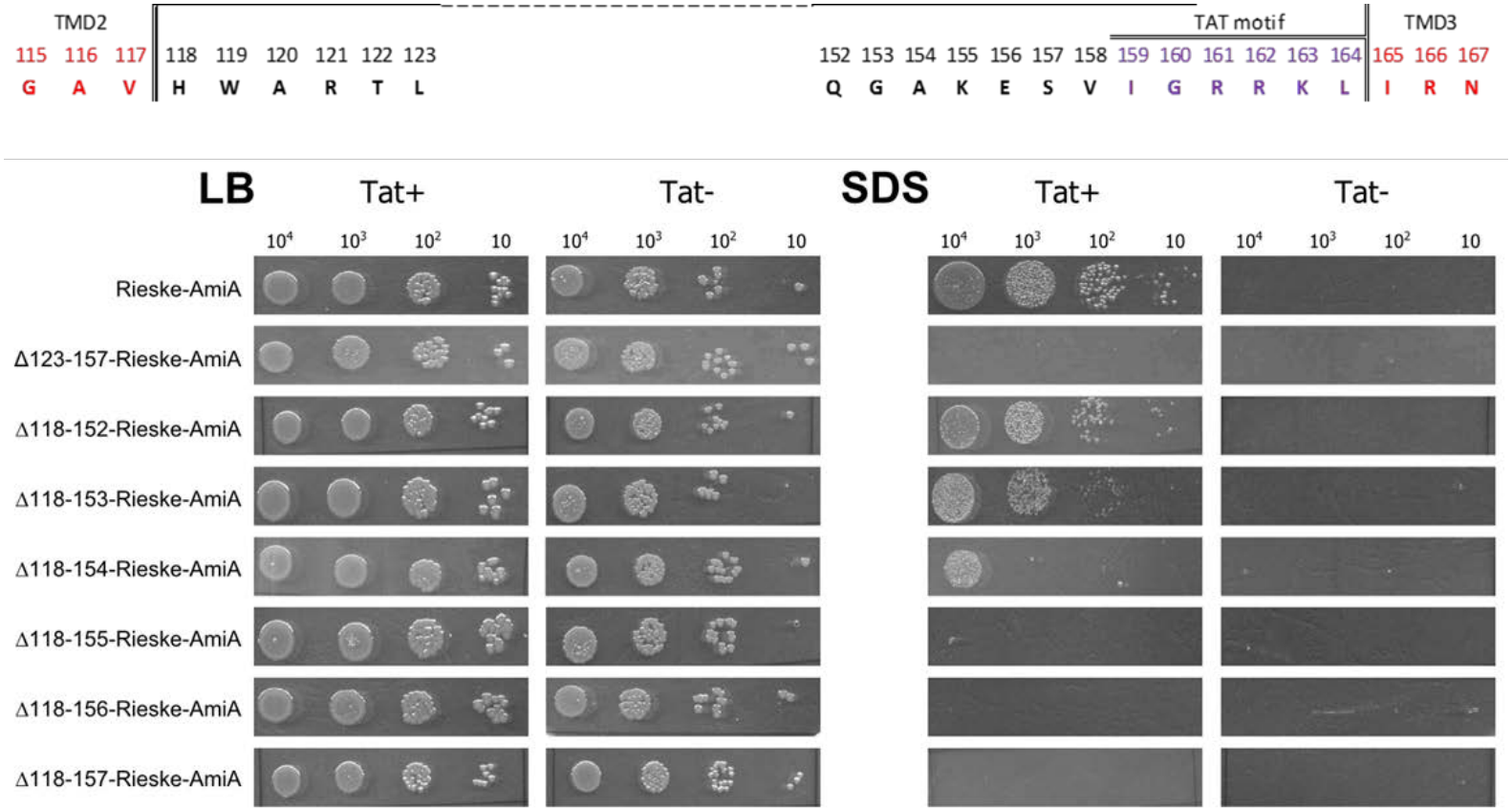


Figure 5.3 RieskeAmiA variants that give reduced or no Tat translocation.
Spot tests of strains MCDSSAC (*tat*$^{+}$ strain) and MCDSSACΔtat (*tat*$^{-}$ strain) harbouring an empty vector (pSUPROM), and the pSUPROM vector encoding Rieske-AmiA and truncation fusion variants (Δ123-157-Rieske-AmiA, Δ118-152-Rieske-AmiA, Δ118-153-Rieske-AmiA, Δ118-154-Rieske-AmiA, Δ118-155-Rieske-AmiA, Δ118-156-Rieske-AmiA and Δ118-157-Rieske-AmiA). Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto LB solid agar and LB solid agar containing 1% (w/v) SDS. LB agar plates were incubated at 37°C for 16 h.
Above is the remainder of the conserved loop region for *S. coelicolor* Rieske for these truncation variants.

5.3.1.4 Significant Rieske-AmiA fusion truncation variants are not recognised by the Tat system

It was seen above that the truncated variants Δ123-157-Rieske-AmiA and Δ118-157-Rieske-AmiA supported no growth on SDS, consistent with a complete inability of the Tat system to recognise these variants. To explore this further, additional variants were generated, where the truncation of Δ118-152 was sequentially increased one residue at a time to Δ118-157, generating the new constructs Δ118-153-, Δ118-154-, Δ118-155- and Δ118-156-Rieske-AmiA.

As seen in Figure 5.3, Δ118-152-Rieske-AmiA shows 'wild-type' growth, whereas the additional deletion of one residue (Δ118-153-Rieske-AmiA) resulted a slight decrease in growth, when expressed in MCDSSAC (*tat*$^+$) strain and tested on SDS-containing media. Thus some very faint growth was seen when 100 cells were spotted but no growth when 10 cells were spotted, seen in Figure 5.3. The Δ118-154-Rieske-AmiA resulted in a further decrease in growth as MCDSSAC harbouring this variant struggled to grow when 10 000 cells were spotted on SDS-containing media. The truncation variants Δ118-155-Rieske-AmiA and Δ118-156-Rieske-AmiA supported no growth of MCDSSAC on SDS-containing media, alongside the previously mentioned Δ123-157-Rieske-AmiA and Δ118-157-Rieske-AmiA (Figure 5.3).

The lack of growth on SDS-containing media for cells harbouring these variants indicates that there is no translocation by the Tat machinery. However, this might not be due to a direct effect of the mutation on recognition by the Tat system, for example as it could arise due to a lack of protein stability. To determine whether stability of the fusion protein was affected by the truncations, membranes from the *tat*$^+$ (MCDSSAC) strain harbouring the variants were analysed by Western blotting with an anti-Rieske antiserum. Representative results are observed in Figure 5.4A, which gives an example of a Western blot of the truncated fusion variants. It is apparent from the blot that a size shift is seen for the truncation variants due to shortening of the protein.

For a quantitative analysis, samples, collected in triplicate and analysed by Western blot, were quantified by densitometry analysis where Rieske-associated signals were normalised to the BamA-associated signals (which served as the loading control). The results were subsequently expressed as percentage of the normalised signal obtained for ‘wild-type’ Rieske-AmiA, as seen in Figure 5.4B. The results indicate that although when some variants are expressed in a *tat*$^{+}$ strain and tested for growth on SDS-containing media there is little Tat-dependent translocation, all fusion proteins are stable and located in the membrane with a considerable amount of protein produced when compared to ‘wild type’ Rieske-AmiA. It can be concluded that the lack of growth seen with some of these variants is due to the truncation of the loop region directly affecting the recognition or translocation of TMD3 and AmiA by the Tat machinery.

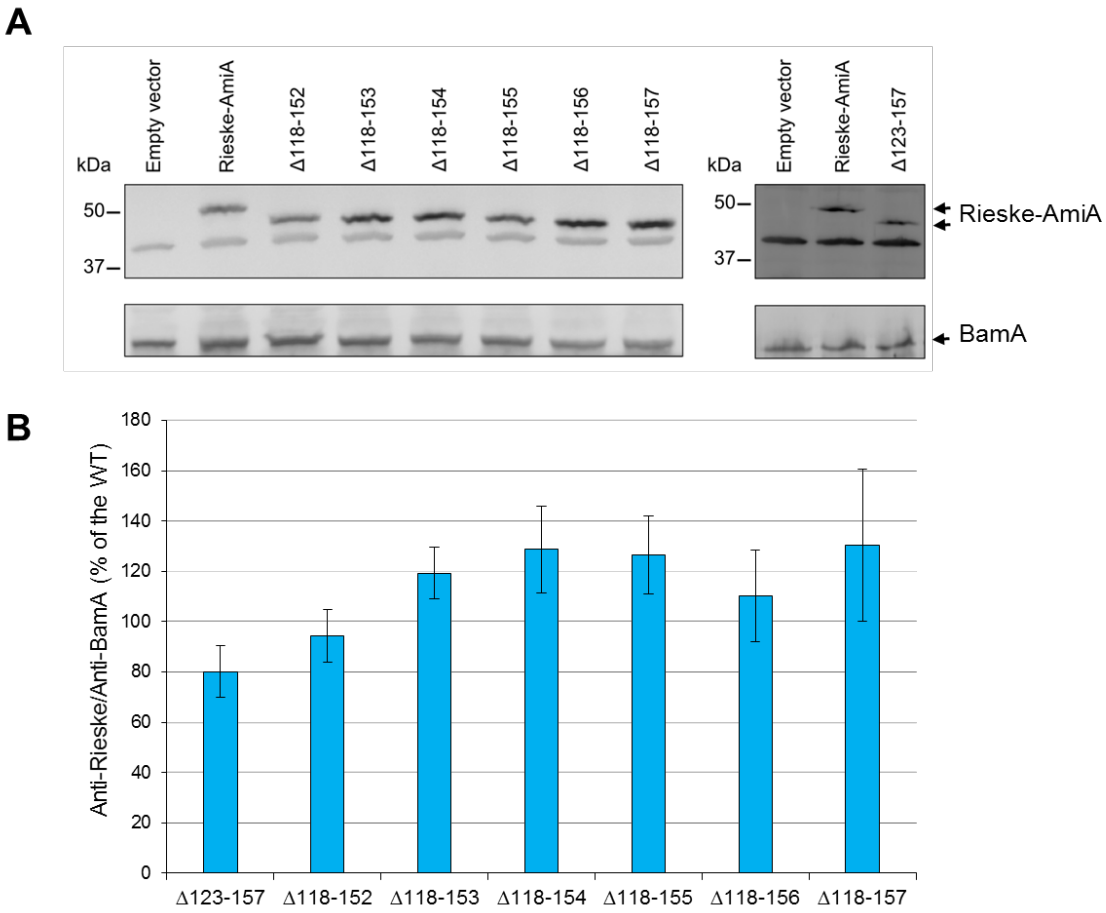


Figure 5.4 Rieske-AmiA protein is still produced from all truncation variants tested
A) Crude membrane extracts of MCDSSAC (*tat*$^+$) harbouring the pSUPROM vector alone (Empty vector) or the pSUPROM vector encoding Rieske-AmiA or loop truncation variants Δ123-157-Rieske-AmiA, Δ118-152-Rieske-AmiA, Δ118-153-Rieske-AmiA, Δ118-154-Rieske-AmiA, Δ118-155-Rieske-AmiA, Δ118-156-Rieske-AmiA and Δ118-157-Rieske-AmiA were separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Rieske peptide antibody. As a loading control, the levels of the OM protein BamA were detected using an anti-BamA antiserum. Western blots used can be found in Figure 8.2.
B) Western blots were quantified by densitometry analysis and Rieske-associated signals were normalised with BamA-associated signals. The results were expressed as percentage of the normalised signal obtained for Rieske-AmiA and the s.e.m is indicated (error bars).

It was seen above that cells producing Δ118-154-Rieske-AmiA (a truncation of 37 residues) struggled to grow on SDS-containing media whereas the same strain producing Δ118-153-Rieske-AmiA only showed a slight decrease in growth. In comparison, the removal of 38 residues from the cytoplasmic loop (Δ118-155-Rieske-AmiA) abolished translocation through the Tat machinery. These results indicate that reducing the cytoplasmic loop region to a length of 6 residues (Δ118-154) or less causes a significant decrease in Tat translocation.

A hypothesis for this loss of Tat translocation is that the cytoplasmic loop region is not long enough to allow exposure of the Tat motif to the translocation machinery. The essential recognition sequence for the Tat machinery is based on a consensus sequence; S**RR**xFLK. This demonstrates that at least one residue, usually a serine, is required before the twin-arginines. In *E. coli* Tat substrates two or more residues are usually present in the N-terminus of the signal peptide prior to the 'serine' position of the Tat motif (Tullman-Ercek *et al.*, 2007). However, the Rieske protein is tethered to the membrane through the TMD1 and TMD2 insertion by the Sec machinery. This would potentially make it more difficult for the Tat machinery to recognise the Tat motif as the N-terminus of the signal sequence is not free and will probably lie very close to the membrane surface, or even within the phospholipid head groups.

5.3.2 Tryptophan at residue 119 is part of TMD2, decreasing the length of the cytoplasmic loop

It was seen from the results presented in Figure 5.3 that seven amino acid residues is the minimum length of the cytoplasmic loop region required for efficient Tat translocation of TMD3. However, it was also seen that the variant Δ123-157-Rieske-AmiA did not support growth on SDS-containing media despite seemingly having eight residues present in the loop region. This Δ123-157-Rieske-AmiA variant has a 35 residue truncation, as intriguingly does its counterpart variant, Δ118-152-Rieske-AmiA, which in contrast did support growth on SDS. To investigate why these two variants differed, the

Δ123-157-Rieske-AmiA variant was lengthened by addition of one residue at a time, to give Δ124-157- and Δ125-157-Rieske-AmiA followed by expression in the *tat*$^{+}$ (MCDSSAC) and *tat*$^{-}$ (MCDSSACΔtat) strains and tested for growth on SDS-containing media.

As Figure 5.5 showed, a Rieske-AmiA truncation lacking residues 124-157 (34 aa truncation) behaved similarly to the Δ123-157-Rieske-AmiA variant and could not support growth on SDS-containing media when expressed in the *tat*$^{+}$ strain, whereas a Rieske-AmiA truncation lacking residues 125-157 had 'wild-type' levels of Tat translocation of AmiA into the periplasm as seen by the growth on SDS-containing media. Thus examination of this truncation series would indicate that the minimum loop length required for Tat translocation is ten residues.

To try and reconcile these two contrasting conclusions (i.e. that the minimum loop length for Tat recognition is seven residues when the truncations started from amino acid 118, versus ten residues when truncations started further along the loop) the loop sequence remaining in these variants was examined more closely. It was noted that all of the truncations starting from 118 were lacking a tryptophan residue which is located N-terminal of the cytoplasmic loop (W119), this residue was present in the truncations starting at residues further along the loop region. Tryptophan has an amphipathic nature and so has a preference for the water-membrane interface. Analysis of membrane protein structures has indicated that aromatic residues, like tryptophan and tyrosine, are frequently located in the interface between the TMDs and the membrane (Hessa *et al.*, 2005). Tryptophan has been found to interact with the polar lipid head groups which attracts it to the membrane, with support from the hydrophobic effect which drives tryptophan out of the water phase (Yau *et al.*, 1998, de Jesus & Allen, 2013). Together, this enables tryptophan to serve as a membrane anchor, adding structural constraint to the movement of TMDs and enhancing stability with interfacial interactions (Granseth *et al.*, 2005, Yau *et al.*, 1998). If tryptophan is located within the loop region then it orientates with its side chain towards the membrane enabling it to be buried within the hydrophobic

bilayer, this facilitates the pulling of the loop into the membrane allowing tryptophan to extend the helix by a residue (Granseth *et al.*, 2005, Braun & von Heijne, 1999).

Taking this knowledge into account, it is likely that W119 and consequently H118 are both located at the interface of TMD2 with the membrane, and should probably be considered to be part of TMD2 rather than residing in the loop region. By making this adjustment, the truncations 123-157, 124-157 and 125-157 should be considered to have loop lengths of six, seven and eight residues, respectively, rather than eight, nine and ten as was assumed previously. To support this proposition, a further set of truncation variants was generated which included residues from both ends of the loop region. HWAKES-Rieske-AmiA contains the H118, W119 and A120 from the N-terminal region of the cytoplasmic loop and the K155, E156 and S157 from the C-terminal, this encompasses a deletion of residues Δ121-154. This variant produces a loop length of 9 residues including H118 and W119 which can, therefore, be considered as a 7 residue loop. The other mutations were generated in a similar fashion and decrease this loop region to 6 residues (HWKES, Δ120-154) and 5 residues (HWES, Δ120-155). As seen in Figure 5.5, all these variants struggle to grow on SDS-containing media.

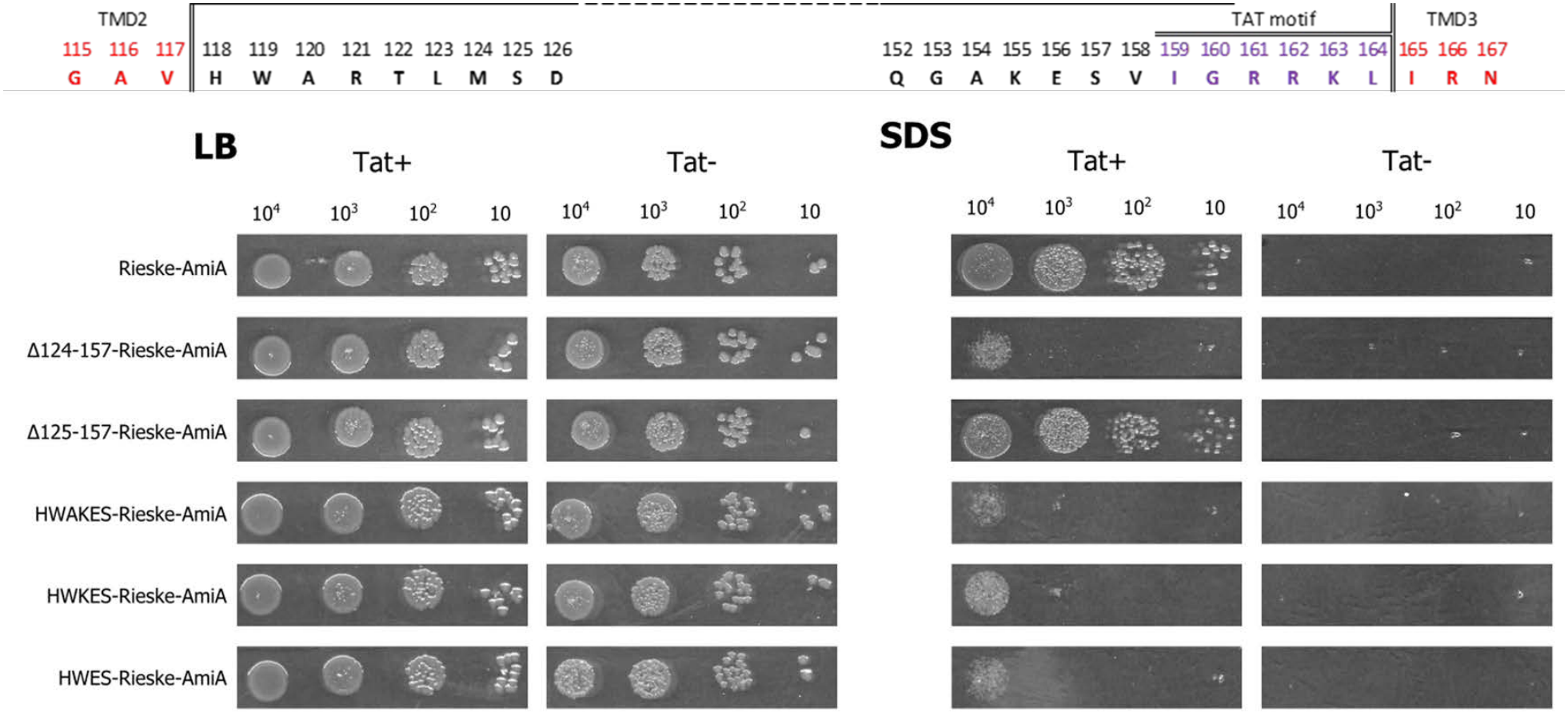


Figure 5.5 The truncated Rieske-AmiA W119-containing variants are not recognised by the Tat system.
Spot tests of strains MCDSSAC (*tat*$^+$ strain) and MCDSSACΔtat (*tat*$^-$ strain) harbouring an empty vector (pSUPROM), and the pSUPROM vector encoding Rieske-AmiA and truncation fusion variants (Δ124-157-Rieske-AmiA, Δ125-157-Rieske-AmiA, HWAKES-Rieske-AmiA, HWKES-Rieske-AmiA and HWES-Rieske-AmiA). Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto LB solid agar and LB solid agar containing 1% (w/v) SDS. LB agar plates were incubated at 37°C for 16 h.
Above is the remainder of the conserved loop region for *S. coelicolor* Rieske for these truncation variants.

To ensure that these variants were stably produced, membranes from the *tat*$^+$ strain (MCDSSAC) harbouring these constructs were analysed by Western blotting with an anti-Rieske antiserum. The results are observed in Figure 5.6 which shows that all the variants produce protein that is stable and located in the membrane. Therefore, the lack of growth on SDS-containing media from some of these variants is due to the truncation of the loop region affecting the recognition of the Tat motif by the Tat machinery.

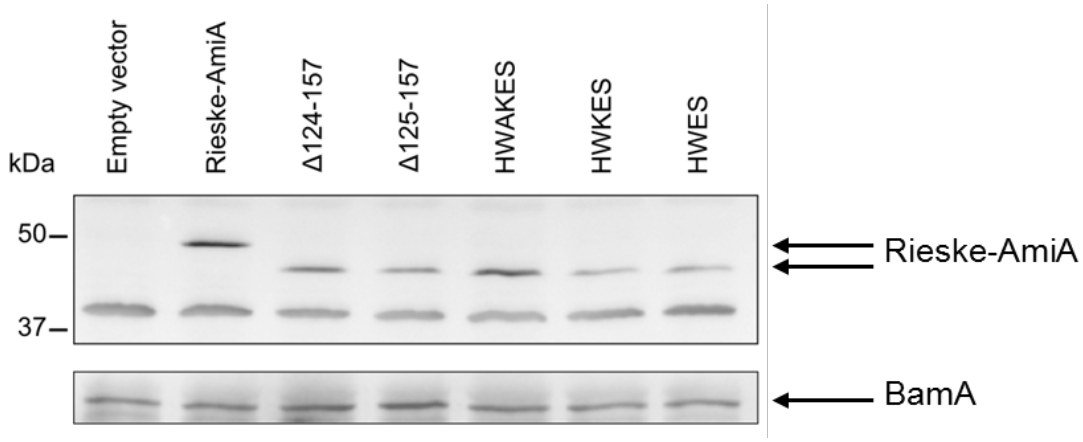


Figure 5.6 The truncated Rieske-AmiA W119-containing variants are stably produced.
Crude membrane extracts of MCDSSAC (*tat*$^+$) harbouring the pSUPROM vector alone (Empty vector) or the pSUPROM vector encoding Rieske-AmiA or loop truncation variants Δ124-157-Rieske-AmiA, Δ123-157-Rieske-AmiA, HWAKES-Rieske-AmiA, HWKES-Rieske-AmiA and HWES-Rieske-AmiA were separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Rieske peptide antibody. As a loading control the levels of the OM protein BamA were detected using an anti-BamA antiserum.

Thus it can be concluded that a minimum length of 7-8 amino acid residues is necessary between the end of TMD2 and the twin-arginine motif to allow the Tat system to recognise TMD3 as a Tat signal sequence, and that conserved sequence features in the loop region, with the exception of the twin-arginine motif, are not directly required for interaction of the substrate with the Tat pathway.

5.4 Discussion

5.4.1 Most of the cytoplasmic loop region is not important for the recognition of the Rieske polypeptide by the Tat machinery

Keller *et al*. (2012) predicted that the cytoplasmic loop region between TMD1 and TMD2 contained all the information required for both the Sec machinery to release the Rieske polypeptide and the Tat machinery to recognise the tethered protein as a Tat substrate. Chapter 4 demonstrated that there are no features within this region required by the Sec machinery to release the partially inserted Rieske protein, although it probably does contain topological signals that are necessary for the correct Sec-dependent insertion of TMD2. To test whether any features in the loop region are important for Tat recognition, mutations of the conserved residues in this loop region were generated in the construct Rieske-AmiA. It was found that large truncations of the loop region had no significant effect on the translocation of Rieske-AmiA (Table 5.1). Therefore, aside from the Tat motif, there are no specific features in the cytoplasmic loop that aid the recognition of Rieske.

The question that must arise as a result of these findings is why is there such length and sequence conservation in the Actinobacterial Rieske TMD2-TMD3 loop regions? Having ruled out a role in protein translocation, it seems likely that the conserved features of this cytoplasmic loop are important for other interactions. In this thesis, the Rieske protein was truncated to remove its cofactor-containing domain, and was analysed outside of the context of its native host. This simplified the study of its assembly into the membrane, by removing any interacting partner proteins (which are shown in cartoon form in Figure 5.7), chaperones and any interactions with the iron-sulphur cluster machinery. In the native organism, the assembly of Rieske is much more complex, and indeed the Tat machinery must not recognise Rieske immediately after its release by the Sec pathway because the iron-sufur cluster cofactor must be inserted, and the protein must be allowed to fold, before it can be translocated across the membrane (Figure 5.8 panel C). Currently nothing is known about how this is achieved for Rieske, however, for other

cofactor-containing Tat substrate proteins, system-specific, proof-reading chaperones bind to Tat signal sequences of their partner proteins to prevent transport before cofactor insertion has occurred (Jack *et al.*, 2004, Maillard *et al.*, 2007, Hatzixanthis *et al.*, 2005, Schubert *et al.*, 2007). It is not known whether a similar proof-reading chaperone exists for Actinobacterial Rieske proteins, however, these system-specific chaperones are usually encoded in operons with the Tat substrate protein they interact with, and there is no candidate ORF close to or within the *S. coelicolor* cytochrome bc_1 coding region. Nonetheless, co-ordinating cofactor-insertion into Rieske prior to translocation could be one important role mediated by the conserved loop region.

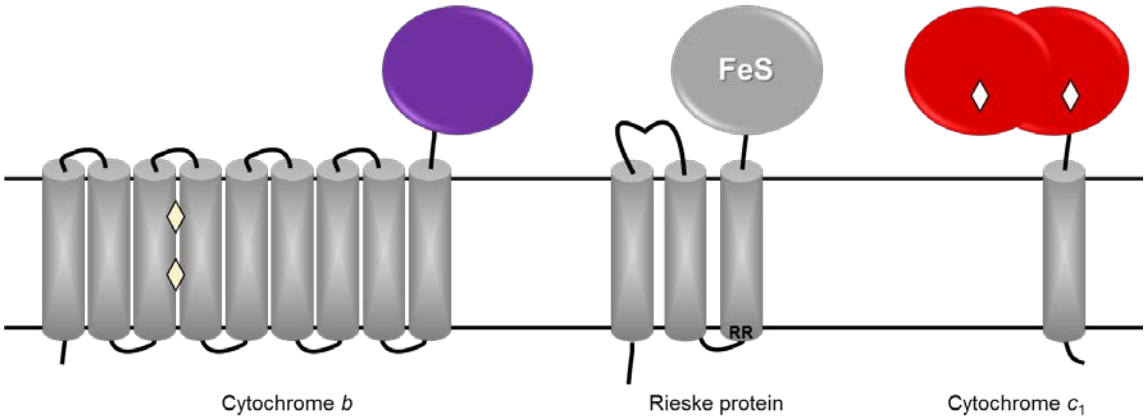
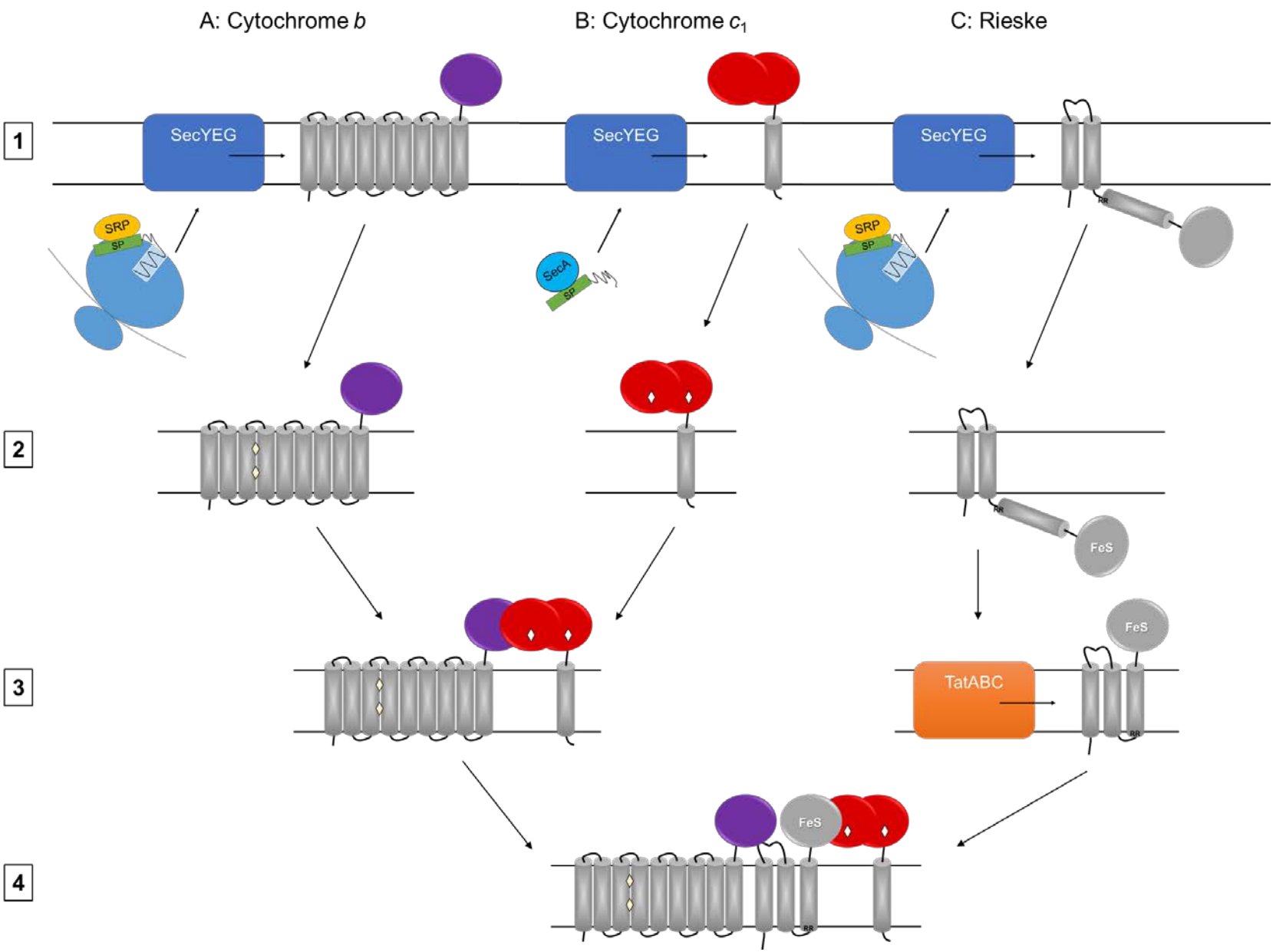


Figure 5.7 Schematic representation of the *S. coelicolor* cytochrome bc_1 complex catalytic components.
A representation of the three subunits of the cytochrome bc_1 complex from *S. coelicolor*. The cofactors are indicated; FeS for iron-sulfur cluster, yellow diamond for a *b*-type heme and white diamond for a *c*-type heme.

A second important role that the loop could serve might be in mediating interactions with partner proteins in the cytochrome bc_1 complex. Prokaryotic bc_1 complexes contain three essential catalytic subunits; cytochrome *b*, cytochrome c_1 and Rieske, seen in Figure 5.7. Cytochrome *b*, in most bacteria, contains 8 TMD, whereas in Actinobacteria this is predicted to be 9 TMD. This protein utilises the SRP/Sec pathway for insertion into the cytoplasmic membrane (Figure 5.8, panel A). Two heme *b* molecules are inserted into the apoprotein following its membrane assembly, the mechanism for this insertion is unknown. Cytochrome c_1, in Actinobacteria, is a membrane-anchored di-heme (Niebisch & Bott, 2003) instead of containing a single heme which interacts with a soluble

cytochrome *c* shuttle (Xia *et al.*, 1997). Its insertion into the membrane is Sec-dependent, presumably mediated by SecA (Figure 5.8, panel B). Once inserted the covalent attachment of heme *c* by the system I cytochrome *c* biogenesis machinery occurs at the extracytoplasmic side of the membrane (Thony-Meyer & Kunzler, 1997). This is an essential step for the folding of the cytochrome and hence generation of the complex. Cytochromes *b* and c_1 have been found to form a protease resistant primary complex before the final addition of the Rieske protein (Davidson *et al.*, 1992). Rieske is usually a single TMD protein with a C-terminal globular FeS cofactor-containing domain, however, in Actinobacteria this protein has an additional 2 N-terminal TMDs (Xia *et al.*, 1997, Kurisu *et al.*, 2003, Keller *et al.*, 2012). The interaction of Rieske with its partner proteins is dependent on the insertion of its cofactor (Figure 5.8). Hopkins *et al.* (2014) found that, in *S. coelicolor,* even when a key component of the Tat system is absent (Δ*tatC*) and the Rieske protein is mislocalised, a stable complex with cytochrome bc_1 components forms. Therefore, it can be proposed that the interactions with the cytochrome *b* and cytochrome c_1 components might in part be mediated by the cytoplasmic loop region.

Cytochrome bc_1 complexes are obligate dimers, with the bacterial Rieske protein contributing significantly to dimer stability by domain swapping; its, usually single, TMD obliquely spans the membrane in one monomer while its extrinsic FeS containing domain is located in the other monomer (Kurisu *et al.*, 2003). However, Actinobacterial Rieske proteins contain additional TMDs which could also contribute to the stability of the complex, with the features of the cytoplasmic loop region potentially involved in interactions to facilitate this stability. This is especially important as other components of the Actinobacterial cytochrome bc_1 complex also show unusual features when compared to the well-characterised complexes from bovine heart mitochondria and *Paracoccus dentrificans* (Baker *et al.*, 1998, Lenaz & Genova, 2012), for example, the protein encoded by *qcrB* (cytochrome *b*, Figure 5.7) likewise has an additional TMD compared to the mitochondrial homologue (Xia *et al.*, 1997).


A: Cytochrome b
B: Cytochrome c1
C: Rieske
1
2
3
4
SecYEG
SecYEG
SecYEG
SRP
SP
SecA
SP
SRP
SP
RR
RR
FeS
FeS
TatABC
RR
FeS
RR

5.4.2 The minimum cytoplasmic loop length for recognition of TMD3 by the Tat pathway is 7 residues

Although there are no specific features in the Rieske protein cytoplasmic loop region (aside from the twin-arginine motif) that are required for recognition of TMD3 by the Tat pathway, it was discovered that there is a specific minimum length requirement for the loop to allow recognition and translocation by the Tat machinery. As seen in Figure 5.3, efficient translocation through the Tat machinery for these variants of Rieske-AmiA requires at least 7 residues immediately before the twin-arginines of the Tat motif.

As discussed in Chapter 4, it is probable that TMD3 remains in the membrane as a re-entrant loop due to the Sec machinery attempting to insert it into the membrane before the recognition of Sec-avoidance positive charges (a schematic is shown in Figure 4.19B). If Rieske has a very short cytoplasmic loop region, the formation of a re-entrant loop along with the tethering to the membrane by the already inserted TMDs would make it problematic for the Tat machinery to recognise the Tat motif which could be embedded in the cytoplasmic membrane. This would explain the requirement for a minimum number of residues within the loop that enable the Tat motif to be accessible for recognition by the Tat machinery while tethered to the membrane.

These results suggest that the Tat machinery is capable of recognising the Tat motif when it is at least partially embedded in the membrane, further supporting the hypothesis that at least a fraction of Tat precursors interact with the membrane as an early stage in Tat translocation.

Figure 5.8 Schematic of *S. coelicolor* cytochrome bc_1 complex biogenesis.
1) Translocation of individual components into the cytoplasmic membrane by the Sec machinery, mediated by SRP or SecA binding to the signal peptide (SP). 2) Insertion of cofactors; *b*-type heme (yellow diamond), *c*-type heme (white diamond) and iron-sulfur cluster (FeS) 3) Interactions between cytochrome *b* and c_1 and the final insertion step of Rieske by the Tat machinery 4) Model of assembled cytochrome bc_1 complex.

6 Discussion

The Rieske proteins of *S. coelicolor*, and from the work in this thesis *M. tuberculosis*, are the only known proteins to be inserted into the cytoplasmic membrane by the concerted action of both the Sec and the Tat machineries (Keller *et al.*, 2012, Hopkins *et al.*, 2014). The work in this thesis has shown that there are no specific features of the cytoplasmic loop region preceding the twin-arginine motif which facilitate its release from the Sec machinery or its recognition by the Tat machinery. Instead it was found that it is the topology determinants for insertion by the Sec machinery that are important in the release of the polypeptide. The combination of the low hydrophobicity of the final TMD and the positive charges located C-terminal to this TMD, enables the Sec machinery to release the partially inserted Rieske protein leaving it membrane anchored. This then allows the cofactor machinery to insert the FeS cluster prior to the Tat machinery recognising the Tat motif and completing the integration process.

The cooperation of Sec and Tat in the insertion of membrane proteins is a complex process which necessitates interplay between two tightly regulated machineries. As previously discussed, the Rieske protein studied here displays extra TMDs compared to most Rieske proteins yet fulfils a similar cellular role. The presence of extra TMDs likely increases the stability of the protein once inserted as well as the overall stability of the cytochrome bc_1 complex as interactions with the partner proteins are mediated *via* the TMDs (Hopkins *et al.*, 2014, Niebisch & Bott, 2001, Sone *et al.*, 2001).

An immediate question that arises from this research is how beneficial is it to the cell to synthesise such complex membrane proteins? The answer to this is not clear. However, it should be noted that there are many examples of multisubunit cofactor-containing proteins with membrane anchors that are synthesised as separate polypeptides rather than a single dual-targeted protein, two possible examples are the *E. coli* isoenzymes hydrogenase 1 and 2, and the *Salmonella enterica* tetrathionate reductase, TtrABC (James *et al.*, 2013).

6.1 Identification of additional dual-targeted membrane proteins

The study by Keller *et al.* (2012) along with work described in this thesis leads to some interesting further questions, such as how widespread across the domains of life is this dual-insertion mode? Is it conserved or is it only found in Actinobacteria? Are there further dual targeted proteins? And if so, is the mechanism of assembly of such proteins conserved?

To identify potential dual-targeted candidates, a stringent set of criteria need to be met. Firstly, the membrane protein would be required to have a Tat motif N-terminal to the final TMD and contain a C-terminal domain that binds a cofactor of the type associated with 'standard' Tat substrates. In addition, an even number of TMDs must be located prior to the Tat motif for the final TMD to be located at the cytoplasmic side of the membrane to allow interaction with the Tat machinery. To identify potential candidates meeting these criteria, a bioinformatics search was undertaken in collaboration with Dr Govind Chandra (John Innes Centre, Norwich; data not shown) where all publicly available bacterial and archaeal protein sequences were interrogated using TatFind (version 1.4 (Rose *et al.*, 2002)) in conjunction with TMHMM (Server v. 2.0, CBS) to select potential dual-targeted candidates. The list was then further sorted manually to identify any potential cofactor-containing domains.

The original search yielded 1712 non-redundant archaeal (58) or bacterial (1654) proteins which displayed an overall odd number of TMDs with a potential Tat motif located just before the last TMD. The results were classified according to the presence of predicted functional domains using EggNOG (Server v. 4.5 (Huerta-Cepas *et al.*, 2016)) and manually analysed for cofactor-containing domains using Pfam 29.0 (Finn *et al.*, 2016). Most candidates were predicted not to contain any cofactor and did not show strong conservation of the twin-arginine motif across closely related species, making them unlikely to be true dual-targeted proteins. However, from this analysis, two further candidates were identified.

The first candidate is a predicted polyferredoxin protein which appears to harbour several 4Fe-4S binding domains, including one at the *trans* side of the membrane and the predicted topology is shown in Figure 6.1B. Homologues of this protein have been found in several different species, five of which are listed in Figure 6.1A and their sequences are aligned in Figure 6.2.

A

Gene	Protein ID	Domain	Phylum	genus/species
MldDRAFT_4510	Q1NSB0	Bacteria	Proteobacteria	*delta proteobacterium MLMS-1*
AMJ46_09695	A0A0S7XC05	Bacteria	Fibrobacteres	*Latescibacteria bacterium DG_63*
AMK71_10745	A0A0S8AJ89	Bacteria	Nitrospirae	*Nitrospira bacterium SG8_35_4*
AMJ88_08315	A0A0S8I6X2	Bacteria	Chloroflexi	*Anaerolineae bacterium SM23_ 63*
MCP_0080	D1YUN0	Archaea	Euryarchaeota	*Methanocella paludicola*

B

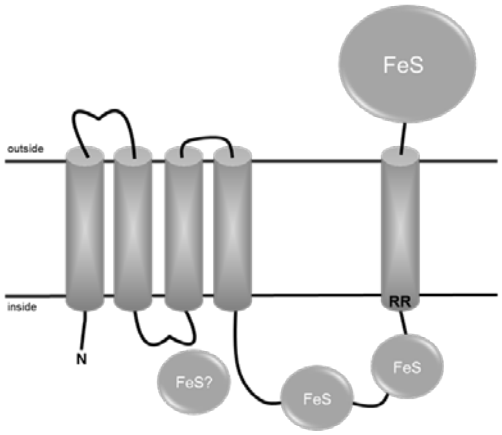


Figure 6.1 Selection of predicted polyferredoxin dual-targeted candidates.
A) A table of five of the predicted dual-targeted polyferredoxin proteins with their taxonomy. B) Cartoon representation of the topology of predicted 5TMD polyferredoxin dual-targeted candidates. The twin-arginine motif is indicated by RR and iron-sulfur binding domains by FeS.

Ferredoxin proteins are involved in numerous biological functions, generally linked to energy conversion/metabolism (Johnson *et al.*, 2005). FeS cluster-containing proteins are frequently found within complexes of the respiratory chain and the export of the FeS domain is generally Tat-dependent (Berks, 1996). 4Fe-4S polyferredoxins are often referred to as bacterial-type ferredoxins. They are involved in a number of different biological functions varying from electron transfer, substrate binding and activation, Fe or cluster storage and regulation of gene expression or enzyme activity. Proteins with

multiple FeS clusters are likely due to the evolution of bacterial-type ferredoxins by intrasequence gene duplication, transposition or fusion events.

There are no clear functions assigned to any of these predicted dual-targeted candidates and deducing the function from neighbouring genes in the operons that encode these proteins is not trivial as mostly the available genome sequences are only partial and partner proteins are not known. However the presence of FeS clusters at the cytoplasmic and periplasmic sides of the membrane suggest that the protein could carry out electron transfer reactions from one side of the membrane to the other.

It is possible that the insertion of the C-terminal FeS cluster into these proteins before export into the cytoplasm is part of an internal quality control mechanism. The twin-arginines are unlikely to be accessible by the Tat system before the cluster is inserted, allowing the cofactor insertion to control the recognition of this region and the final translocation of the protein. This is an important step in the mechanism of insertion of dual-targeted proteins and would prevent translocation before the protein is fully matured.

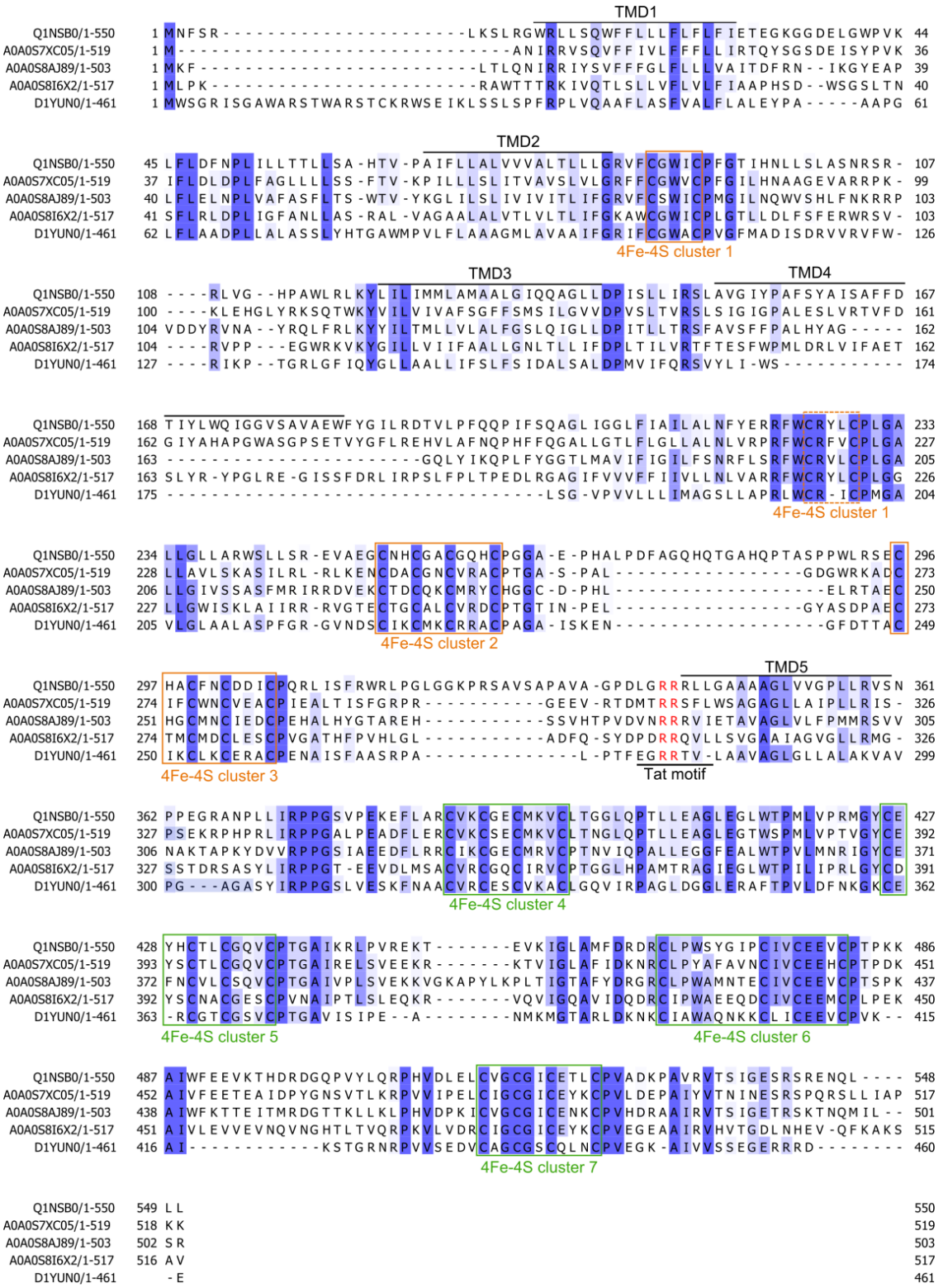


Figure 6.2 A sequence alignment of predicted polyferredoxin dual-targeted candidates.
The sequence alignment was performed with Jalview using TcoffeeWS (default settings) of the predicted polyferredoxin candidates listed in Figure 6.1A. The Tat motif is underlined with the twin-arginines in red and the residues involved in Iron–sulphur ligation predicted using MotifFinder are boxed (orange for 4Fe-4S before the Tat motif and green for 4Fe-4S in the periplasmic C-terminal extremity). Conserved residues are colour-coded according to Jalview, using a conservation threshold of 30. The position of the TMDs was predicted with TMHMM and TMPred and is indicated with black line for the template sequence of PFD (Q1NSB0).

The second dual-targeted candidate protein contains a C-terminal molybdopterin binding domain (MoCo). Prokaryotic molybdenum-associated proteins are also Tat-dependent with a complicated insertion process occurring in the cytoplasm before export (Iobbi-Nivol & Leimkuhler, 2013, Berks, 1996). Homologues of the predicted MoCo-containing protein from *S. coelicolor* (SCO3746) were found in 86 different species of bacteria and archaea, some of which are shown in Figure 6.3A; 76 of these are in Actinobacteria. These proteins are predicted to have 5TMDs prior to the C-terminal-MoCo domain. They also display a predicted cytochrome b_{561} binding domain at their N-terminal region, as shown in Figure 6.3B. Despite being reported as absent from prokaryotes and fungi (Verelst & Asard, 2003), a family of cytochrome b_{561} annotated proteins in prokaryotes (Murakami *et al.*, 1986) and more recently fungal homologues (Tsubaki et al., 2005) have been described. Whereas the function in ascorbate recycling/iron absorption of cytochrome b_{561} in eukaryotes has been extensively studied (Lu *et al.*, 2014), roles for these proteins in bacteria remain elusive. Figure 6.4 shows an alignment of the proteins in Figure 6.3A, indicating the cofactor binding domains and the Tat motif.

A

Gene	Protein ID	Domain	Phylum	genus/species	PF00174		PF01292	
					Position	i-Evalue	Position	i-Evalue
SCO3746	Q9L0V8	Bacteria	Actinobacteria	*Streptomyces coelicolor A3(2)*	290..425	2e-22	36-117	0.25
HFX_0636	I3R2A0	Archaea	Euryarchaeota	*Haloferax mediterranei*	212..348	1e-29	-	-
Rv0218	P96400	Bacteria	Actinobacteria	*Mycobacterium tuberculosis H37Rv*	310..441	1.7e-18	25-121	0.0044
Rcas_4179	A7NRL4	Bacteria	Chloroflexi	*Roseiflexus castenholzii*	198..332	5.9e-27	88-143	0.02
Aaci_0290	C8WRE4	Bacteria	Firmicutes	*Alicyclobacillus acidocaldarius*	215..360	1.6e-39	10-138	0.013

B

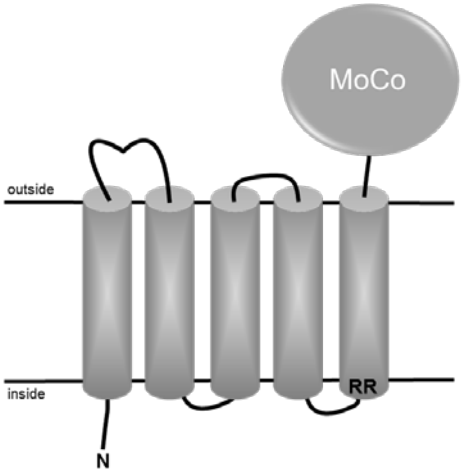


Figure 6.3 A selection of predicted molybdopterin dual-targeted proteins.
A) A table of some of the predicted candidate proteins and their taxonomy. Position and i-Evalue of domains is indicated. PF00174: Oxidoreductase molybdopterin binding domain (Pfam ID: Oxidored_molyb) and PF01292: Prokaryotic cytochrome b_{561} (Pfam ID: Ni_hydr_CYTB). B) Cartoon representation of the topology of predicted 5TMD molybdopterin dual-targeted proteins. The twin-arginine motif is indicated by RR and the molybdopterin-binding domain by MoCo.

As with the predicted polyferredoxin candidates, these proteins also do not have any assigned function. However, a number of molybdoenzymes are also sulphite-oxidizing enzymes (SOE) with a function as sulphite oxidases or sulphite dehydrogenases. Bacterial SOEs, containing a MoCo domain are exported by the Tat system into the periplasm (Brokx *et al.*, 2005) where they have been shown to catalyse sulphite oxidation. There are a subgroup of bacterial heme-containing SOEs where the Tat-dependent MoCo subunit interacts with a membrane-intrinsic heme-containing subunit which is Sec-dependent (Kappler, 2011). The domains of these two subunits, YedY and YedZ, mimics the predicted MoCo dual-targeted proteins which indicates a possible function as a sulphite dehydrogenase, as in these proteins the electrons are transferred from the cytochrome to the MoCo domain. The identification of such proteins also raises the possibility that heme-MoCo dual-targeted proteins arose during the course of evolution from separate polypeptides. Alternatively, the ancestoral protein may

have been a dual-targeted fusion of the heme and MoCo domains that subsequently separated in some organisms.

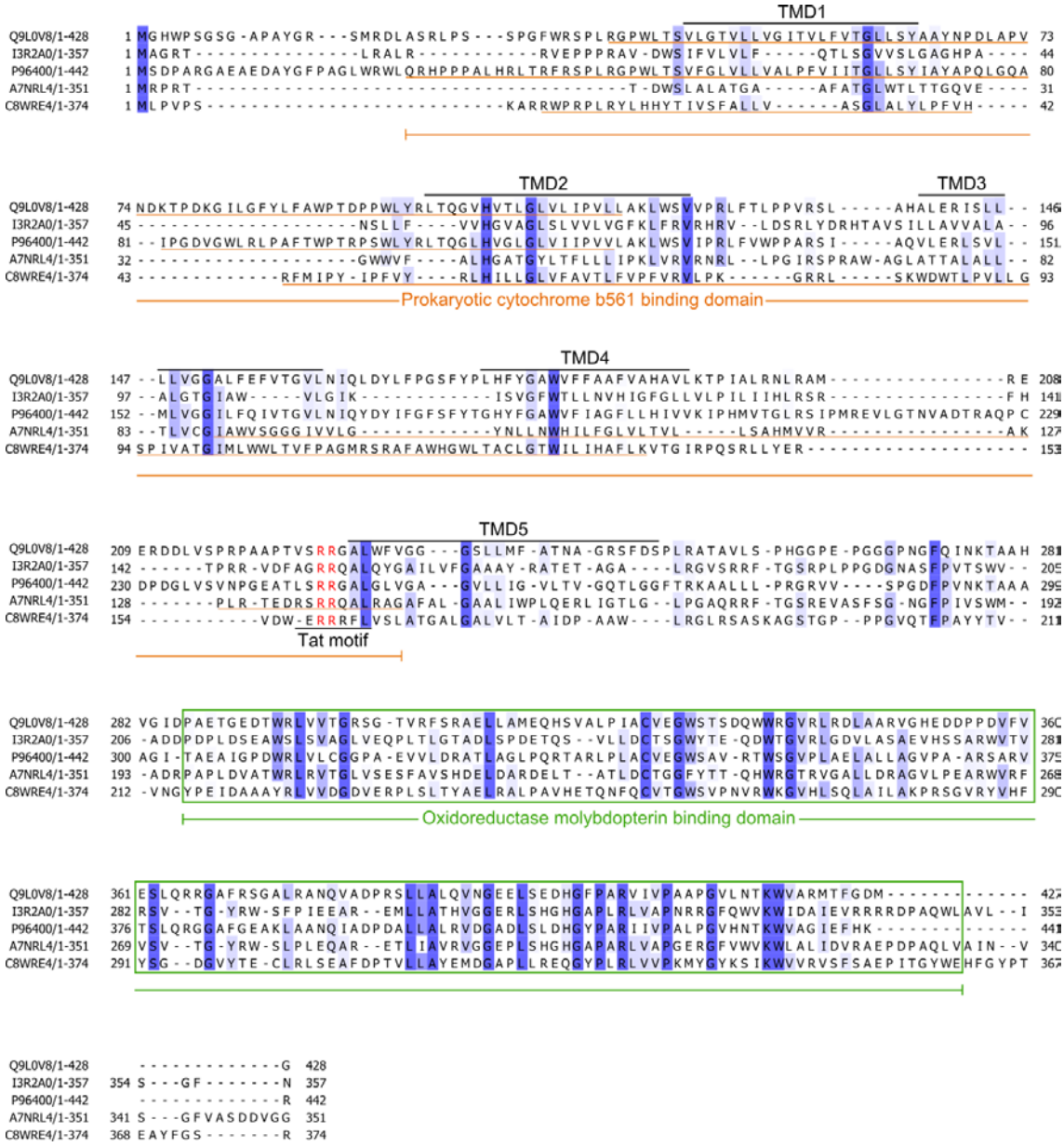


Figure 6.4 Sequence alignment of predicted MoCo dual-targeted proteins.
The sequence alignment was performed with Jalview using TcoffeeWS (default settings) using the proteins indicated in Figure 6.3A. The Tat motif is underlined with the twin-arginine in red and the residues involved in the molybdopterin binding domain predicted using MotifFinder are boxed (Green). The predicted cytochrome b_{561}binding domain is underlined in orange. Conserved residues are colour-coded according to Jalview, using a conservation threshold of 30. The position of the TMDs was predicted with TMHMM and TMPred and is indicated with black line for the template sequence of Sco3746 (Q9L0V8).

The bioinformatics search also highlighted around a hundred QcrA (Rieske) proteins with 3TMDs, suggesting that Rieske is the most common dual-targeted protein. Surprisingly a 5TMD Rieske protein (KSE_30950) was also found in the Actinobacterial *Kitasatospora setae* strain. This organism is known to produce setamycin (bafilomycin B1) which has antimicrobial activity, and it is morphologically similar to the genus *Streptomyces* (Ichikawa *et al*., 2010). KSE_30950 is a putative Rieske protein predicted to contain 5TMDs with a Tat motif directly before the final TMD which is followed by an FeS cluster domain. It is interesting to speculate why there are two further TMDs within this organism, perhaps due to a duplication of the TMDs.

Actinobacteria had a higher number of predicted dual-targeted candidates than any other phylum. As *Streptomyces* have also been predicted to have a large number of Tat substrates (Schaerlaekens *et al.*, 2004, Widdick *et al.*, 2008, Joshi *et al.*, 2010), this indicates that the Tat pathway might have a more important role in these bacteria.

6.2 Is the mechanism of dual-targeting of *S. coelicolor* Rieske conserved?

The mechanism of dual-targeting established in this work uses the topology requirements for Sec substrates and it was concluded that the combination of hydrophobicity and positive charges cause the Sec machinery to release the translocating polypeptide allowing the rest of the insertion to occur by the Tat pathway. Analysis of the most likely predicted dual-targeted substrates (Sco3746, PFD and 5TMD Rieske) found that these proteins, as with *S. coelicolor* Rieske, have lower hydrophobicity of the final TMD, calculated by a positive ΔG value (Table 6.1) and have one or more positive charges located directly after the final TMD. Although dual-targeting will have to be confirmed experimentally it is very likely that these three candidates are inserted into the cytoplasmic membrane by the co-operation of the Sec and Tat machineries.

Table 6.1 Predicted ΔG values for TMDs of predicted dual-targeted proteins.

Strain	Protein	Number of TMDs	Sequence	Predicted ΔG
S. coelicolor	Rieske	3	IRNTMLGALTLVPLSGVVLLR	1.215
Kitasatospora setae	Rieske	5	LDGGIAAVALGGAGLLSGSLTA	1.1312
S. coelicolor	Sco3746	5	RGALWFVGGGSLLMFATNAGRSF	2.166
***deltaproteobacterium* MLMS-1**	PFD	5	RLLGAAAAGLVVGPLLRVSN	1.419

The presence of additional TMDs in these proven and potential dual-targeted proteins is likely due to evolution. Gene duplication is known as an important mechanism for adapting to charges in environment (Coenye *et al.*, 2005) and it is possible that TMDs were gained in this process alongside the mechanism which evolved to cope with the assembly of these proteins. It also possible that they have become evolutionary redundant and this mechanism is only retained by the organisms where the larger more hydrophobic proteins are essential. This is probably due to diverse locations, dormant life cycles and survival in extreme stress conditions.

The evolution of this dual-targeted system can only be speculated on until more experimentally confirmed substrates have been identified and a clear evolutionary pathway investigated. The bioinformatics search has found a large number of potential candidates, however, as there are proteins that have been reported to bind cofactors with an unusual motif that differs to the standard conserved motifs (Rao *et al.*, 2011), these proteins would not have been included in the final output. Further work on uncharacterised proteins and novel binding motifs may lead to the identification of a larger number of substrates that require this dual insertion mode in the future.

7 Bibliography

Akimaru, J., S. Matsuyama, H. Tokuda & S. Mizushima, (1991) Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from *Escherichia coli*. *Proc Natl Acad Sci U S A* **88**: 6545-6549.

Alami, M., I. Luke, S. Deitermann, G. Eisner, H.G. Koch, J. Brunner & M. Muller, (2003) Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. *Mol Cell* **12**: 937-946.

Aldridge, C., E. Spence, M.A. Kirkilionis, L. Frigerio & C. Robinson, (2008) Tat-dependent targeting of Rieske iron-sulphur proteins to both the plasma and thylakoid membranes in the cyanobacterium *Synechocystis* PCC6803. *Mol Microbiol* **70**: 140-150.

Andersson, H., E. Bakker & G. von Heijne, (1992) Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in *Escherichia coli*. *J Biol Chem* **267**: 1491-1495.

Andersson, H. & G. von Heijne, (1994) Membrane protein topology: effects of Δmu H+ on the translocation of charged residues explain the 'positive inside' rule. *Embo J* **13**: 2267-2272.

Angelini, S., S. Deitermann & H.G. Koch, (2005) FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. *Embo Rep* **6**: 476-481.

Asai, T., Y. Shinoda, T. Nohara, T. Yoshihisa & T. Endo, (1999) Sec-dependent pathway and ΔpH-dependent pathway do not share a common translocation pore in thylakoidal protein transport. *J Biol Chem* **274**: 20075-20078.

Bachmann, J., B. Bauer, K. Zwicker, B. Ludwig & O. Anderka, (2006) The Rieske protein from *Paracoccus denitrificans* is inserted into the cytoplasmic membrane by the twin-arginine translocase. *Febs J* **273**: 4817-4830.

Bageshwar, U.K., N. Whitaker, F.C. Liang & S.M. Musser, (2009) Interconvertibility of lipid- and translocon-bound forms of the bacterial Tat precursor pre-SufI. *Mol Microbiol* **74**: 209-226.

Baker, S.C., S.J. Ferguson, B. Ludwig, M.D. Page, O.M. Richter & R.J. van Spanning, (1998) Molecular genetics of the genus *Paracoccus*: metabolically versatile bacteria with bioenergetic flexibility. *Microbiology and molecular biology reviews : MMBR* **62**: 1046-1078.

Barka, E.A., P. Vatsa, L. Sanchez, N. Gaveau-Vaillant, C. Jacquard, H.P. Klenk, C. Clement, Y. Ouhdouch & G.P. van Wezel, (2016) Taxonomy, Physiology, and Natural Products of *Actinobacteria*. *Microbiology and molecular biology reviews : MMBR* **80**: 1-43.

Barkocy-Gallagher, G.A. & P.J. Bassford, Jr., (1992) Synthesis of precursor maltose-binding protein with proline in the +1 position of the cleavage site interferes with the activity of *Escherichia coli* signal peptidase I *in vivo*. *J Biol Chem* **267**: 1231-1238.

Barnett, J.P., R.T. Eijlander, O.P. Kuipers & C. Robinson, (2008) A minimal Tat system from a gram-positive organism: a bifunctional TatA subunit participates in discrete TatAC and TatA complexes. *J Biol Chem* **283**: 2534-2542.

Barrett, C.M., J.E. Mathers & C. Robinson, (2003) Identification of key regions within the *Escherichia coli* TatAB subunits. *Febs Lett* **537**: 42-46.

Barrett, C.M. & C. Robinson, (2005) Evidence for interactions between domains of TatA and TatB from mutagenesis of the TatABC subunits of the twin-arginine translocase. *Febs J* **272**: 2261-2275.

Bartolome, B., Y. Jubete, E. Martinez & F. de la Cruz, (1991) Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. *Gene* **102**: 75-78.

Bassilana, M. & W. Wickner, (1993) Purified *Escherichia coli* preprotein translocase catalyzes multiple cycles of precursor protein translocation. *Biochemistry* **32**: 2626-2630.

Bauer, J., M.J. Fritsch, T. Palmer & G. Unden, (2011) Topology and accessibility of the transmembrane helices and the sensory site in the bifunctional transporter DcuB of *Escherichia coli*. *Biochemistry* **50**: 5925-5938.

Bechtluft, P., R.G. van Leeuwen, M. Tyreman, D. Tomkiewicz, N. Nouwen, H.L. Tepper, A.J. Driessen & S.J. Tans, (2007) Direct observation of chaperone-induced changes in a protein folding pathway. *Science* **318**: 1458-1461.

Beck, K., G. Eisner, D. Trescher, R.E. Dalbey, J. Brunner & M. Muller, (2001) YidC, an assembly site for polytopic *Escherichia coli* membrane proteins located in immediate proximity to the SecYE translocon and lipids. *Embo Rep* **2**: 709-714.

Beck, K., L.F. Wu, J. Brunner & M. Muller, (2000) Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. *Embo J* **19**: 134-143.

Beckmann, R., C.M. Spahn, N. Eswar, J. Helmers, P.A. Penczek, A. Sali, J. Frank & G. Blobel, (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* **107**: 361-372.

Bendtsen, J.D., H. Nielsen, D. Widdick, T. Palmer & S. Brunak, (2005) Prediction of twin-arginine signal peptides. *Bmc Bioinformatics* **6**: 167.

Bentley, S.D., K.F. Chater, A.M. Cerdeno-Tarraga, G.L. Challis, N.R. Thomson, K.D. James, D.E. Harris, M.A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C.W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B.G. Barrell, J. Parkhill & D.A. Hopwood, (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141-147.

Berks, B.C., (1996) A common export pathway for proteins binding complex redox cofactors? *Mol Microbiol* **22**: 393-404.

Berks, B.C., (2015) The twin-arginine protein translocation pathway. *Annu Rev Biochem* **84**: 843-864.

Berks, B.C., S.M. Lea & P.J. Stansfeld, (2014) Structural biology of Tat protein transport. *Current opinion in structural biology* **27**: 32-37.

Berks, B.C., T. Palmer & F. Sargent, (2005) Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current opinion in microbiology* **8**: 174-181.

Berks, B.C., F. Sargent & T. Palmer, (2000) The Tat protein export pathway. *Mol Microbiol* **35**: 260-274.

Bernhardt, T.G. & P.A. de Boer, (2003) The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported *via* the twin-arginine transport pathway. *Mol Microbiol* **48**: 1171-1182.

Berry, E.A., M. Guergova-Kuras, L.S. Huang & A.R. Crofts, (2000) Structure and function of cytochrome *bc* complexes. *Annu Rev Biochem* **69**: 1005-1075.

Bieker-Brady, K. & T.J. Silhavy, (1992) Suppressor analysis suggests a multistep, cyclic mechanism for protein secretion in *Escherichia coli*. *Embo J* **11**: 3165-3174.

Bieker, K.L., G.J. Phillips & T.J. Silhavy, (1990) The *sec* and *prl* genes of *Escherichia coli*. *Journal of bioenergetics and biomembranes* **22**: 291-310.

Birnboim, H.C. & J. Doly, (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**: 1513-1523.

Blattner, F.R., G. Plunkett, 3rd, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau & Y. Shao, (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1462.

Blaudeck, N., P. Kreutzenbeck, R. Freudl & G.A. Sprenger, (2003) Genetic analysis of pathway specificity during posttranslational protein translocation across the *Escherichia coli* plasma membrane. *J Bacteriol* **185**: 2811-2819.

Blobel, G., (1980) Intracellular protein topogenesis. *Proc Natl Acad Sci U S A* **77**: 1496-1500.

Blummel, A.S., L.A. Haag, E. Eimer, M. Muller & J. Frobel, (2015) Initial assembly steps of a translocase for folded proteins. *Nat Commun* **6**: 7234.

Bogdanov, M., W. Dowhan & H. Vitrac, (2014) Lipids and topological rules governing membrane protein assembly. *Biochim Biophys Acta* **1843**: 1475-1488.

Bogsch, E., S. Brink & C. Robinson, (1997) Pathway specificity for a ΔpH-dependent precursor thylakoid lumen protein is governed by a 'Sec-avoidance' motif in the transfer peptide and a 'Sec-incompatible' mature protein. *The EMBO journal* **16**: 3851-3859.

Bogsch, E.G., F. Sargent, N.R. Stanley, B.C. Berks, C. Robinson & T. Palmer, (1998) An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *J Biol Chem* **273**: 18003-18006.

Bolhuis, A., J.E. Mathers, J.D. Thomas, C.M. Barrett & C. Robinson, (2001) TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. *J Biol Chem* **276**: 20213-20219.

Botelho, S.C., M. Osterberg, A.S. Reichert, K. Yamano, P. Bjorkholm, T. Endo, G. von Heijne & H. Kim, (2011) TIM23-mediated insertion of transmembrane α-helices into the mitochondrial inner membrane. *Embo J* **30**: 1003-1011.

Braun, P. & G. von Heijne, (1999) The aromatic residues Trp and Phe have different effects on the positioning of a transmembrane helix in the microsomal membrane. *Biochemistry* **38**: 9778-9782.

Brokx, S.J., R.A. Rothery, G. Zhang, D.P. Ng & J.H. Weiner, (2005) Characterization of an *Escherichia coli* sulfite oxidase homologue reveals the role of a conserved active site cysteine in assembly and function. *Biochemistry* **44**: 10339-10348.

Broome-Smith, J.K. & B.G. Spratt, (1986) A vector for the construction of translational fusions to TEM β-lactamase and the analysis of protein export signals and membrane protein topology. *Gene* **49**: 341-349.

Broome-Smith, J.K., M. Tadayyon & Y. Zhang, (1990) β-lactamase as a probe of membrane protein assembly and protein export. *Mol Microbiol* **4**: 1637-1644.

Brundage, L., J.P. Hendrick, E. Schiebel, A.J. Driessen & W. Wickner, (1990) The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62**: 649-657.

Bruser, T. & C. Sanders, (2003) An alternative model of the twin arginine translocation system. *Microbiological research* **158**: 7-17.

Buchanan, G., E. de Leeuw, N.R. Stanley, M. Wexler, B.C. Berks, F. Sargent & T. Palmer, (2002) Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis. *Mol Microbiol* **43**: 1457-1470.

Buchanan, G., F. Sargent, B.C. Berks & T. Palmer, (2001) A genetic screen for suppressors of *Escherichia coli* Tat signal peptide mutations establishes a critical role for the second arginine within the twin-arginine motif. *Arch Microbiol* **177**: 107-112.

Buck, T.M., J. Wagner, S. Grund & W.R. Skach, (2007) A novel tripartite motif involved in aquaporin topogenesis, monomer folding and tetramerization. *Nature structural & molecular biology* **14**: 762-769.

Caldelari, I., T. Palmer & F. Sargent, (2008) *Escherichia coli tat* mutant strains are able to transport maltose in the absence of an active *malE* gene. *Arch Microbiol* **189**: 597-604.

Casadaban, M.J. & S.N. Cohen, (1979) Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: *in vivo* probe for transcriptional control sequences. *P Natl Acad Sci USA* **76**: 4530-4533.

Cerdeno-Tarraga, A.M., A. Efstratiou, L.G. Dover, M.T. Holden, M. Pallen, S.D. Bentley, G.S. Besra, C. Churcher, K.D. James, A. De Zoysa, T. Chillingworth, A. Cronin, L. Dowd, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Moule, M.A. Quail, E. Rabbinowitsch, K.M. Rutherford, N.R. Thomson, L. Unwin, S. Whitehead, B.G.

Barrell & J. Parkhill, (2003) The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res* **31**: 6516-6523.

Chaddock, A.M., A. Mant, I. Karnauchov, S. Brink, R.G. Herrmann, R.B. Klosgen & C. Robinson, (1995) A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the ΔpH-dependent thylakoidal protein translocase. *Embo J* **14**: 2715-2722.

Challis, G.L. & D.A. Hopwood, (2003) Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces species*. *Proc Natl Acad Sci U S A* **100 Suppl 2**: 14555-14561.

Chang, C.Y., L. Hobley, R. Till, M. Capeness, M. Kanna, W. Burtt, P. Jagtap, S. Aizawa & R.E. Sockett, (2011) The *Bdellovibrio bacteriovorus* twin-arginine transport system has roles in predatory and prey-independent growth. *Microbiology* **157**: 3079-3093.

Chatzi, K.E., M.F. Sardis, A. Economou & S. Karamanou, (2014) SecA-mediated targeting and translocation of secretory proteins. *Bba-Mol Cell Res* **1843**: 1466-1474.

Chen, Y., R. Soman, S.K. Shanmugam, A. Kuhn & R.E. Dalbey, (2014) The role of the strictly conserved positively charged residue differs among the Gram-positive, Gram-negative, and chloroplast YidC homologs. *J Biol Chem* **289**: 35656-35667.

Clark, S.A. & S.M. Theg, (1997) A folded protein can be transported across the chloroplast envelope and thylakoid membranes. *Mol Biol Cell* **8**: 923-934.

Cline, K. & M. McCaffery, (2007) Evidence for a dynamic and transient pathway through the TAT protein transport machinery. *The EMBO journal* **26**: 3039-3049.

Cline, K. & H. Mori, (2001) Thylakoid ΔpH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. *Journal of Cell Biology* **154**: 719-729.

Coenye, T., D. Gevers, Y. Van de Peer, P. Vandamme & J. Swings, (2005) Towards a prokaryotic genomic taxonomy. *FEMS microbiology reviews* **29**: 147-167.

Cole, S.T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.E. Sulston, K. Taylor, S. Whitehead & B.G. Barrell, (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Collier, D.N., V.A. Bankaitis, J.B. Weiss & P.J. Bassford, Jr., (1988) The antifolding activity of SecB promotes the export of the *E. coli* maltose-binding protein. *Cell* **53**: 273-283.

Cooley, J.W., (2013) Protein conformational changes involved in the cytochrome bc_1 complex catalytic cycle. *Biochim Biophys Acta* **1827**: 1340-1345.

Cristobal, S., J.W. de Gier, H. Nielsen & G. von Heijne, (1999) Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. *The EMBO journal* **18**: 2982-2990.

Crofts, A.R., (2004) The cytochrome bc_1 complex: function in the context of structure. *Annual review of physiology* **66**: 689-733.

Cronan, J.E., (2003) Bacterial membrane lipids: where do we stand? *Annu Rev Microbiol* **57**: 203-224.

Dabney-Smith, C. & K. Cline, (2009) Clustering of C-terminal stromal domains of Tha4 homo-oligomers during translocation by the Tat protein transport system. *Mol Biol Cell* **20**: 2060-2069.

Dabney-Smith, C., H. Mori & K. Cline, (2006) Oligomers of Tha4 organize at the thylakoid Tat translocase during protein transport. *J Biol Chem* **281**: 5476-5483.

Dalbey, R.E., A. Kuhn, L. Zhu & D. Kiefer, (2014) The membrane insertase YidC. *Bba-Mol Cell Res* **1843**: 1489-1496.

Danese, P.N. & T.J. Silhavy, (1998) Targeting and assembly of periplasmic and outer-membrane proteins in *Escherichia coli*. *Annual review of genetics* **32**: 59-94.

Darrouzet, E., J.W. Cooley & F. Daldal, (2004) The Cytochrome bc_1 Complex and its Homologue the *b* (6) *f* Complex: Similarities and Differences. *Photosynthesis research* **79**: 25-44.

Davidson, E., T. Ohnishi, M. Tokito & F. Daldal, (1992) *Rhodobacter capsulatus* mutants lacking the Rieske FeS protein form a stable cytochrome bc_1 subcomplex with an intact quinone reduction site. *Biochemistry* **31**: 3351-3358.

De Buck, E., L. Vranckx, E. Meyen, L. Maes, L. Vandersmissen, J. Anne & E. Lannnertyn, (2007) The twin-arginine translocation pathway is necessary for correct membrane insertion of the Rieske Fe/S protein in *Legionella pneumophila*. *Febs Lett* **581**: 259-264.

De Gier, J.W., Q.A. Valent, G. Von Heijne & J. Luirink, (1997) The *E. coli* SRP: preferences of a targeting factor. *Febs Lett* **408**: 1-4.

de Jesus, A.J. & T.W. Allen, (2013) The role of tryptophan side chains in membrane protein anchoring and hydrophobic mismatch. *Biochim Biophys Acta* **1828**: 864-876.

de Leeuw, E., T. Granjon, I. Porcelli, M. Alami, S.B. Carr, M. Muller, F. Sargent, T. Palmer & B.C. Berks, (2002) Oligomeric properties and signal peptide binding by *Escherichia coli* Tat protein transport complexes. *J Mol Biol* **322**: 1135-1146.

De Leeuw, E., I. Porcelli, F. Sargent, T. Palmer & B.C. Berks, (2001) Membrane interactions and self-association of the TatA and TatB components of the twin-arginine translocation pathway. *Febs Lett* **506**: 143-148.

de Muinck, E.J., K. Lagesen, J.E. Afset, X. Didelot, K.S. Ronningen, K. Rudi, N.C. Stenseth & P. Trosvik, (2013) Comparisons of infant *Escherichia coli* isolates link genomic profiles with adaptation to the ecological niche. *BMC genomics* **14**: 81.

Dekker, C., B. de Kruijff & P. Gros, (2003) Crystal structure of SecB from *Escherichia coli*. *Journal of structural biology* **144**: 313-319.

DeLisa, M.P., P. Samuelson, T. Palmer & G. Georgiou, (2002) Genetic analysis of the twin arginine translocator secretion pathway in bacteria. *J Biol Chem* **277**: 29825-29831.

DeLisa, M.P., D. Tullman & G. Georgiou, (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *P Natl Acad Sci USA* **100**: 6115-6120.

Demirci, E., T. Junne, S. Baday, S. Berneche & M. Spiess, (2013) Functional asymmetry within the Sec61p translocon. *Proc Natl Acad Sci U S A* **110**: 18856-18861.

Devaraneni, P.K., B. Conti, Y. Matsumura, Z. Yang, A.E. Johnson & W.R. Skach, (2011) Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. *Cell* **146**: 134-147.

Dilks, K., M.I. Gimenez & M. Pohlschroder, (2005) Genetic and biochemical analysis of the twin-arginine translocation pathway in halophilic archaea. *J Bacteriol* **187**: 8104-8113.

Dilks, K., R.W. Rose, E. Hartmann & M. Pohlschroder, (2003) Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey. *J Bacteriol* **185**: 1478-1483.

Dow, J.M., F. Gabel, F. Sargent & T. Palmer, (2013) Characterization of a pre-export enzyme-chaperone complex on the twin-arginine transport pathway. *The Biochemical journal* **452**: 57-66.

Dow, J.M., S. Grahl, R. Ward, R. Evans, O. Byron, D.G. Norman, T. Palmer & F. Sargent, (2014) Characterization of a periplasmic nitrate reductase in complex with its biosynthetic chaperone. *Febs J* **281**: 246-260.

Dramsi, S., S. Magnet, S. Davison & M. Arthur, (2008) Covalent attachment of proteins to peptidoglycan. *FEMS microbiology reviews* **32**: 307-320.

du Plessis, D.J., N. Nouwen & A.J. Driessen, (2011) The Sec translocase. *Biochim Biophys Acta* **1808**: 851-865.

Duong, F. & W. Wickner, (1997a) Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *Embo J* **16**: 2756-2768.

Duong, F. & W. Wickner, (1997b) The SecDFyajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *Embo J* **16**: 4871-4879.

Egea, P.F., S.O. Shan, J. Napetschnig, D.F. Savage, P. Walter & R.M. Stroud, (2004) Substrate twinning activates the signal recognition particle and its receptor. *Nature* **427**: 215-221.

Egea, P.F. & R.M. Stroud, (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci U S A* **107**: 17182-17187.

Eimer, E., J. Frobel, A.S. Blummel & M. Muller, (2015) TatE as a Regular Constituent of Bacterial Twin-arginine Protein Translocases. *J Biol Chem* **290**: 29281-29289.

Enquist, K., M. Fransson, C. Boekel, I. Bengtsson, K. Geiger, L. Lang, A. Pettersson, S. Johansson, G. von Heijne & I. Nilsson, (2009) Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J Mol Biol* **387**: 1153-1164.

Esko, J.D., T.L. Doering & C.R.H. Raetz, (2009) Eubacteria and Archaea. In: Essentials of Glycobiology. A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart & M.E. Etzler (eds). Cold Spring Harbor (NY), pp.

Fekkes, P. & A.J. Driessen, (1999) Protein targeting to the bacterial cytoplasmic membrane. *Microbiology and molecular biology reviews : MMBR* **63**: 161-173.

Fekkes, P., C. van der Does & A.J. Driessen, (1997) The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *Embo J* **16**: 6105-6113.

Ferris, F.G. & T.J. Beveridge, (1986) Physicochemical roles of soluble metal cations in the outer membrane of *Escherichia coli* K-12. *Canadian journal of microbiology* **32**: 594-601.

Finn, R.D., P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S.C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G.A. Salazar, J. Tate & A. Bateman, (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279-285.

Frauenfeld, J., J. Gumbart, E.O. Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten & R. Beckmann, (2011) Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nature structural & molecular biology* **18**: 614-621.

Freudl, R., (2013) Leaving home ain't easy: protein export systems in Gram-positive bacteria. *Res Microbiol* **164**: 664-674.

Freymann, D.M., R.J. Keenan, R.M. Stroud & P. Walter, (1997) Structure of the conserved GTPase domain of the signal recognition particle. *Nature* **385**: 361-364.

Frobel, J., P. Rose, F. Lausberg, A.S. Blummel, R. Freudl & M. Muller, (2012) Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. *Nat Commun* **3**.

Frobel, J., P. Rose & M. Muller, (2011) Early contacts between substrate proteins and TatA translocase component in twin-arginine translocation. *J Biol Chem* **286**: 43679-43689.

Fu, L.M. & C.S. Fu-Liu, (2002) Is *Mycobacterium tuberculosis* a closer relative to Gram-positive or Gram-negative bacterial pathogens? *Tuberculosis* **82**: 85-90.

Fujita, H., Y. Kida, M. Hagiwara, F. Morimoto & M. Sakaguchi, (2010) Positive Charges of Translocating Polypeptide Chain Retrieve an Upstream Marginal Hydrophobic Segment from the Endoplasmic Reticulum Lumen to the Translocon. *Mol Biol Cell* **21**: 2045-2056.

Gafvelin, G. & G. Vonheijne, (1994) Topological Frustration in Multispanning *Escherichia-Coli* Inner Membrane-Proteins. *Cell* **77**: 401-412.

Gannon, P.M., P. Li & C.A. Kumamoto, (1989) The mature portion of *Escherichia coli* maltose-binding protein (MBP) determines the dependence of MBP on SecB for export. *J Bacteriol* **171**: 813-818.

Gardel, C., K. Johnson, A. Jacq & J. Beckwith, (1990) The *secD* locus of *E. coli* codes for two membrane proteins required for protein export. *Embo J* **9**: 4205-4206.

Gavel, Y., J. Steppuhn, R. Herrmann & G. von Heijne, (1991) The 'positive-inside rule' applies to thylakoid membrane proteins. *Febs Lett* **282**: 41-46.

Gerard, F. & K. Cline, (2006) Efficient twin arginine translocation (Tat) pathway transport of a precursor protein covalently anchored to its initial cpTatC binding site. *J Biol Chem* **281**: 6130-6135.

Gerard, F. & K. Cline, (2007) The thylakoid proton gradient promotes an advanced stage of signal peptide binding deep within the Tat pathway receptor complex. *J Biol Chem* **282**: 5263-5272.

Getahun, H., A. Matteelli, R.E. Chaisson & M. Raviglione, (2015) Latent *Mycobacterium tuberculosis* infection. *The New England journal of medicine* **372**: 2127-2135.

Goder, V., T. Junne & M. Spiess, (2004) Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol Biol Cell* **15**: 1470-1478.

Goder, V. & M. Spiess, (2003) Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *Embo J* **22**: 3645-3653.

Gogala, M., T. Becker, B. Beatrix, J.P. Armache, C. Barrio-Garcia, O. Berninghausen & R. Beckmann, (2014) Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* **506**: 107-110.

Gohlke, U., L. Pullan, C.A. McDevitt, I. Porcelli, E. de Leeuw, T. Palmer, H.R. Saibil & B.C. Berks, (2005) The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc Natl Acad Sci U S A* **102**: 10482-10486.

Goodfellow, M. & S.T. Williams, (1983) Ecology of actinomycetes. *Annu Rev Microbiol* **37**: 189-216.

Goosens, V.J., C.G. Monteferrante & J.M. van Dijl, (2014) Co-factor insertion and disulfide bond requirements for twin-arginine translocase-dependent export of the *Bacillus subtilis* Rieske protein QcrA. *J Biol Chem* **289**: 13124-13131.

Gouridis, G., S. Karamanou, I. Gelis, C.G. Kalodimos & A. Economou, (2009) Signal peptides are allosteric activators of the protein translocase. *Nature* **462**: 363-367.

Granseth, E., G. von Heijne & A. Elofsson, (2005) A study of the membrane-water interface region of membrane proteins. *J Mol Biol* **346**: 377-385.

Grant, S.G., J. Jessee, F.R. Bloom & D. Hanahan, (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *P Natl Acad Sci USA* **87**: 4645-4649.

Gu, S.Q., F. Peske, H.J. Wieden, M.V. Rodnina & W. Wintermeyer, (2003) The signal recognition particle binds to protein L23 at the peptide exit of the *Escherichia coli* ribosome. *Rna* **9**: 566-573.

Gumbart, J. & K. Schulten, (2008) The roles of pore ring and plug in the SecY protein-conducting channel. *The Journal of general physiology* **132**: 709-719.

Gutierrez-Cirlos, E.B., T. Merbitz-Zahradnik & B.L. Trumpower, (2002) Failure to insert the iron-sulfur cluster into the Rieske iron-sulfur protein impairs both center N and center P of the cytochrome bc_1 complex. *J Biol Chem* **277**: 50703-50709.

Guzman, L.M., D. Belin, M.J. Carson & J. Beckwith, (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *J Bacteriol* **177**: 4121-4130.

Hainzl, T., S. Huang, G. Merilainen, K. Brannstrom & A.E. Sauer-Eriksson, (2011) Structural basis of signal-sequence recognition by the signal recognition particle. *Nature structural & molecular biology* **18**: 389-391.

Halic, M., M. Blau, T. Becker, T. Mielke, M.R. Pool, K. Wild, I. Sinning & R. Beckmann, (2006) Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* **444**: 507-511.

Hanada, M., K.I. Nishiyama, S. Mizushima & H. Tokuda, (1994) Reconstitution of an efficient protein translocation machinery comprising SecA and the three membrane proteins, SecY, SecE, and SecG (p12). *J Biol Chem* **269**: 23625-23631.

Harley, C.A., J.A. Holt, R. Turner & D.J. Tipper, (1998) Transmembrane protein insertion orientation in yeast depends on the charge difference across transmembrane segments, their total hydrophobicity, and its distribution. *J Biol Chem* **273**: 24963-24971.

Hartl, F.U., S. Lecker, E. Schiebel, J.P. Hendrick & W. Wickner, (1990) The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *E. coli* plasma membrane. *Cell* **63**: 269-279.

Hatzixanthis, K., T.A. Clarke, A. Oubrie, D.J. Richardson, R.J. Turner & F. Sargent, (2005) Signal peptide-chaperone interactions on the twin-arginine protein transport pathway. *Proc Natl Acad Sci U S A* **102**: 8460-8465.

Hatzixanthis, K., T. Palmer & F. Sargent, (2003) A subset of bacterial inner membrane proteins integrated by the twin-arginine translocase. *Mol Microbiol* **49**: 1377-1390.

Hedin, L.E., K. Ojemalm, A. Bernsel, A. Hennerdal, K. Illergard, K. Enquist, A. Kauko, S. Cristobal, G. von Heijne, M. Lerch-Bader, I. Nilsson & A. Elofsson, (2010) Membrane Insertion of Marginally Hydrophobic Transmembrane Helices Depends on Sequence Context. *J Mol Biol* **396**: 221-229.

Heidrich, C., M.F. Templin, A. Ursinus, M. Merdanovic, J. Berger, H. Schwarz, M.A. de Pedro & J.V. Holtje, (2001) Involvement of N-acetylmuramyl-$_L$-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* **41**: 167-178.

Heijne, G., (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *The EMBO journal* **5**: 3021-3027.

Hennon, S.W., R. Soman, L. Zhu & R.E. Dalbey, (2015) YidC/Alb3/Oxa1 Family of Insertases. *J Biol Chem* **290**: 14866-14874.

Hessa, T., H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, I. Nilsson, S.H. White & G. von Heijne, (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**: 377-381.

Hessa, T., N.M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S.H. White & G. von Heijne, (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**: 1026-U1022.

Hessa, T., J.H. Reithinger, G. von Heijne & H. Kim, (2009) Analysis of transmembrane helix integration in the endoplasmic reticulum in *S. cerevisiae*. *J Mol Biol* **386**: 1222-1228.

Hicks, M.G., P.A. Lee, G. Georgiou, B.C. Berks & T. Palmer, (2005) Positive selection for loss-of-function *tat* mutations identifies critical residues required for TatA activity. *J Bacteriol* **187**: 2920-2925.

Higy, M., T. Junne & M. Spiess, (2004) Topogenesis of membrane proteins at the endoplasmic reticulum. *Biochemistry* **43**: 12716-12722.

Hinsley, A.P., N.R. Stanley, T. Palmer & B.C. Berks, (2001) A naturally occurring bacterial Tat signal peptide lacking one of the 'invariant' arginine residues of the consensus targeting motif. *Febs Lett* **497**: 45-49.

Hoffmann, C., A. Leis, M. Niederweis, J.M. Plitzko & H. Engelhardt, (2008) Disclosure of the *mycobacterial* outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc Natl Acad Sci U S A* **105**: 3963-3967.

Holzapfel, E., G. Eisner, M. Alami, C.M. Barrett, G. Buchanan, I. Luke, J.M. Betton, C. Robinson, T. Palmer, M. Moser & M. Muller, (2007) The entire N-terminal half of TatC is involved in twin-arginine precursor binding. *Biochemistry* **46**: 2892-2898.

Holzapfel, E., M. Moser, E. Schiltz, T. Ueda, J.M. Betton & M. Muller, (2009) Twin-arginine-dependent translocation of SufI in the absence of cytosolic helper proteins. *Biochemistry* **48**: 5096-5105.

Hopkins, A., G. Buchanan & T. Palmer, (2014) Role of the twin arginine protein transport pathway in the assembly of the *Streptomyces coelicolor* cytochrome bc_1 complex. *J Bacteriol* **196**: 50-59.

Hou, B., S. Frielingsdorf & R.B. Klosgen, (2006) Unassisted membrane insertion as the initial step in ΔpH/Tat-dependent protein transport. *J Mol Biol* **355**: 957-967.

Hu, Y., E. Zhao, H. Li, B. Xia & C. Jin, (2010) Solution NMR structure of the TatA component of the twin-arginine protein transport system from Gram-positive bacterium *Bacillus subtilis*. *Journal of the American Chemical Society* **132**: 15942-15944.

Huerta-Cepas, J., D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M.C. Walter, T. Rattei, D.R. Mende, S. Sunagawa, M. Kuhn, L.J. Jensen, C. von Mering & P. Bork, (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44**: D286-293.

Hunte, C., J. Koepke, C. Lange, T. Rossmanith & H. Michel, (2000) Structure at 2.3 A resolution of the cytochrome bc_1 complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Structure* **8**: 669-684.

Hynds, P.J., D. Robinson & C. Robinson, (1998) The Sec-independent twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane. *J Biol Chem* **273**: 34868-34874.

Ichikawa, N., A. Oguchi, H. Ikeda, J. Ishikawa, S. Kitani, Y. Watanabe, S. Nakamura, Y. Katano, E. Kishi, M. Sasagawa, A. Ankai, S. Fukui, Y. Hashimoto, S. Kamata, M. Otoguro, S. Tanikawa, T. Nihira, S. Horinouchi, Y. Ohnishi, M. Hayakawa, T. Kuzuyama, A. Arisawa, F. Nomoto, H. Miura, Y. Takahashi & N. Fujita, (2010) Genome sequence of *Kitasatospora setae* NBRC 14216T: an evolutionary snapshot of the family *Streptomycetaceae*. *DNA research : an international journal for rapid publication of reports on genes and genomes* **17**: 393-406.

Ignatova, Z., C. Hornle, A. Nurk & V. Kasche, (2002) Unusual signal peptide directs penicillin amidase from *Escherichia coli* to the Tat translocation machinery. *Biochem Biophys Res Commun* **291**: 146-149.

Inouye, S., X. Soberon, T. Franceschini, K. Nakamura, K. Itakura & M. Inouye, (1982) Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. *P Natl Acad Sci USA* **79**: 3438-3441.

Iobbi-Nivol, C. & S. Leimkuhler, (2013) Molybdenum enzymes, their maturation and molybdenum cofactor biosynthesis in *Escherichia coli*. *Biochim Biophys Acta* **1827**: 1086-1101.

Iwata, S., J.W. Lee, K. Okada, J.K. Lee, M. Iwata, B. Rasmussen, T.A. Link, S. Ramaswamy & B.K. Jap, (1998) Complete structure of the 11-subunit bovine mitochondrial cytochrome bc_1 complex. *Science* **281**: 64-71.

Ize, B., F. Gerard & L.F. Wu, (2002a) *In vivo* assessment of the Tat signal peptide specificity in *Escherichia coli*. *Arch Microbiol* **178**: 548-553.

Ize, B., F. Gerard, M. Zhang, A. Chanal, R. Voulhoux, T. Palmer, A. Filloux & L.F. Wu, (2002b) *In vivo* dissection of the Tat translocation pathway in *Escherichia coli*. *J Mol Biol* **317**: 327-335.

Ize, B., N.R. Stanley, G. Buchanan & T. Palmer, (2003) Role of the *Escherichia coli* Tat pathway in outer membrane integrity. *Mol Microbiol* **48**: 1183-1193.

Jack, R.L., G. Buchanan, A. Dubini, K. Hatzixanthis, T. Palmer & F. Sargent, (2004) Coordinating assembly and export of complex bacterial proteins. *The EMBO journal* **23**: 3962-3972.

Jack, R.L., F. Sargent, B.C. Berks, G. Sawers & T. Palmer, (2001) Constitutive expression of *Escherichia coli tat* genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth. *J Bacteriol* **183**: 1801-1804.

Jain, M., C.J. Petzold, M.W. Schelle, M.D. Leavell, J.D. Mougous, C.R. Bertozzi, J.A. Leary & J.S. Cox, (2007) Lipidomics reveals control of *Mycobacterium tuberculosis* virulence lipids *via* metabolic coupling. *Proc Natl Acad Sci U S A* **104**: 5133-5138.

James, M.J., S.J. Coulthurst, T. Palmer & F. Sargent, (2013) Signal peptide etiquette during assembly of a complex respiratory enzyme. *Mol Microbiol* **90**: 400-414.

Janda, C.Y., J. Li, C. Oubridge, H. Hernandez, C.V. Robinson & K. Nagai, (2010) Recognition of a signal peptide by the signal recognition particle. *Nature* **465**: 507-510.

Jiang, F., M. Chen, L. Yi, J.W. de Gier, A. Kuhn & R.E. Dalbey, (2003) Defining the regions of *Escherichia coli* YidC that contribute to activity. *J Biol Chem* **278**: 48965-48972.

Johnson, D.C., D.R. Dean, A.D. Smith & M.K. Johnson, (2005) Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem* **74**: 247-281.

Johnson, T.J. & L.K. Nolan, (2009) Pathogenomics of the virulence plasmids of *Escherichia coli*. *Microbiology and molecular biology reviews : MMBR* **73**: 750-774.

Jormakka, M., S. Tornroth, B. Byrne & S. Iwata, (2002) Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. *Science* **295**: 1863-1868.

Joshi, M.V., S.G. Mann, H. Antelmann, D.A. Widdick, J.K. Fyans, G. Chandra, M.I. Hutchings, I. Toth, M. Hecker, R. Loria & T. Palmer, (2010) The twin arginine protein transport pathway exports multiple virulence proteins in the plant pathogen *Streptomyces scabies*. *Mol Microbiol* **77**: 252-271.

Kamio, Y. & H. Nikaido, (1976) Outer membrane of *Salmonella typhimurium*: accessibility of phospholipid head groups to phospholipase *c* and cyanogen bromide activated dextran in the external medium. *Biochemistry* **15**: 2561-2570.

Kappler, U., (2011) Bacterial sulfite-oxidizing enzymes. *Biochim Biophys Acta* **1807**: 1-10.

Keenan, R.J., D.M. Freymann, P. Walter & R.M. Stroud, (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* **94**: 181-191.

Keller, R., J. de Keyzer, A.J.M. Driessen & T. Palmer, (2012) Co-operation between different targeting pathways during integration of a membrane protein. *Journal of Cell Biology* **199**: 303-315.

Kihara, A., Y. Akiyama & K. Ito, (1995) FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proc Natl Acad Sci U S A* **92**: 4532-4536.

Kim, H., S. Paul, J. Gennity, J. Jennity & M. Inouye, (1994) Reversible topology of a bifunctional transmembrane protein depends upon the charge balance around its transmembrane domain. *Mol Microbiol* **11**: 819-831.

Klenner, C. & A. Kuhn, (2012) Dynamic disulfide scanning of the membrane-inserting Pf3 coat protein reveals multiple YidC substrate contacts. *J Biol Chem* **287**: 3769-3776.

Kneuper, H., B. Maldonado, F. Jager, M. Krehenbrink, G. Buchanan, R. Keller, M. Muller, B.C. Berks & T. Palmer, (2012) Molecular dissection of TatC defines critical regions essential for protein transport and a TatB-TatC contact site. *Mol Microbiol* **85**: 945-961.

Koch, S., M.J. Fritsch, G. Buchanan & T. Palmer, (2012) *Escherichia coli* TatA and TatB proteins have N-out, C-in topology in intact cells. *J Biol Chem* **287**: 14420-14431.

Kramer, G., T. Rauch, W. Rist, S. Vorderwulbecke, H. Patzelt, A. Schulze-Specking, N. Ban, E. Deuerling & B. Bukau, (2002) L23 protein functions as a chaperone docking site on the ribosome. *Nature* **419**: 171-174.

Kreutzenbeck, P., C. Kroger, F. Lausberg, N. Blaudeck, G.A. Sprenger & R. Freudl, (2007) *Escherichia coli* twin arginine (Tat) mutant translocases possessing relaxed signal peptide recognition specificities. *Journal of Biological Chemistry* **282**: 7903-7911.

Kuhn, P., B. Weiche, L. Sturm, E. Sommer, F. Drepper, B. Warscheid, V. Sourjik & H.G. Koch, (2011) The bacterial SRP receptor, SecA and the ribosome use overlapping binding sites on the SecY translocon. *Traffic* **12**: 563-578.

Kumazaki, K., S. Chiba, M. Takemoto, A. Furukawa, K. Nishiyama, Y. Sugano, T. Mori, N. Dohmae, K. Hirata, Y. Nakada-Nakura, A.D. Maturana, Y. Tanaka, H. Mori, Y. Sugita, F. Arisaka, K. Ito, R. Ishitani, T. Tsukazaki & O. Nureki, (2014a) Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**: 516-520.

Kumazaki, K., T. Kishimoto, A. Furukawa, H. Mori, Y. Tanaka, N. Dohmae, R. Ishitani, T. Tsukazaki & O. Nureki, (2014b) Crystal structure of *Escherichia coli* YidC, a membrane protein chaperone and insertase. *Scientific reports* **4**: 7299.

Kumazaki, K., T. Tsukazaki, T. Nishizawa, Y. Tanaka, H.E. Kato, Y. Nakada-Nakura, K. Hirata, Y. Mori, H. Suga, N. Dohmae, R. Ishitani & O. Nureki, (2014c) Crystallization and preliminary X-ray diffraction analysis of YidC, a membrane-protein chaperone and insertase from *Bacillus halodurans*. *Acta crystallographica. Section F, Structural biology communications* **70**: 1056-1060.

Kurisu, G., H. Zhang, J.L. Smith & W.A. Cramer, (2003) Structure of the cytochrome b_6f complex of oxygenic photosynthesis: tuning the cavity. *Science* **302**: 1009-1014.

Laemmli, U.K., (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Laws, J.K. & R.E. Dalbey, (1989) Positive charges in the cytoplasmic domain of *Escherichia coli* leader peptidase prevent an apolar domain from functioning as a signal. *Embo J* **8**: 2095-2099.

Leake, M.C., N.P. Greene, R.M. Godun, T. Granjon, G. Buchanan, S. Chen, R.M. Berry, T. Palmer & B.C. Berks, (2008) Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by *in vivo* single-molecule imaging. *P Natl Acad Sci USA* **105**: 15376-15381.

Lee, H.J. & K.T. Hughes, (2006) Posttranscriptional control of the *Salmonella enterica* flagellar hook protein FlgE. *J Bacteriol* **188**: 3308-3316.

Lee, P.A., G. Buchanan, N.R. Stanley, B.C. Berks & T. Palmer, (2002) Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation. *J Bacteriol* **184**: 5871-5879.

Lehr, U., M. Schutz, P. Oberhettinger, F. Ruiz-Perez, J.W. Donald, T. Palmer, D. Linke, I.R. Henderson & I.B. Autenrieth, (2010) C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of Yersinia enterocolitica are decisive for its recognition and assembly by BamA. *Mol Microbiol* **78**: 932-946.

Leive, L., (1968) Studies on the permeability change produced in coliform bacteria by ethylenediaminetetraacetate. *J Biol Chem* **243**: 2373-2380.

Lenaz, G. & M.L. Genova, (2012) Supramolecular organisation of the mitochondrial respiratory chain: a new challenge for the mechanism and control of oxidative phosphorylation. *Advances in experimental medicine and biology* **748**: 107-144.

Lerch-Bader, M., C. Lundin, H. Kim, I. Nilsson & G. von Heijne, (2008) Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **105**: 4127-4132.

Li, H., P.E. Jacques, M.G. Ghinet, R. Brzezinski & R. Morosoli, (2005) Determining the functionality of putative Tat-dependent signal peptides in *Streptomyces coelicolor* A3(2) by using two different reporter proteins. *Microbiology* **151**: 2189-2198.

Ligon, L.S., J.D. Hayden & M. Braunstein, (2012) The ins and outs of *Mycobacterium tuberculosis* protein export. *Tuberculosis* **92**: 121-132.

Liu, H. & J.H. Naismith, (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**: 91.

Lowry, O.H., N.J. Rosebrough, A.L. Farr & R.J. Randall, (1951) Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**: 265-275.

Lu, P., D. Ma, C. Yan, X. Gong, M. Du & Y. Shi, (2014) Structure and mechanism of a eukaryotic transmembrane ascorbate-dependent oxidoreductase. *Proc Natl Acad Sci U S A* **111**: 1813-1818.

Luirink, J., C.M. ten Hagen-Jongman, C.C. van der Weijden, B. Oudega, S. High, B. Dobberstein & R. Kusters, (1994) An alternative protein targeting pathway in *Escherichia coli*: studies on the role of FtsY. *Embo J* **13**: 2289-2296.

Luke, I., J.I. Handford, T. Palmer & F. Sargent, (2009) Proteolytic processing of *Escherichia coli* twin-arginine signal peptides by LepB. *Arch Microbiol* **191**: 919-925.

Lukjancenko, O., T.M. Wassenaar & D.W. Ussery, (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microbial ecology* **60**: 708-720.

Lundin, C., H. Kim, I. Nilsson, S.H. White & G. von Heijne, (2008) Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N_{in}-C_{out} and N_{out}-C_{in} transmembrane helices. *Proc Natl Acad Sci U S A* **105**: 15702-15707.

Ma, X. & K. Cline, (2010) Multiple precursor proteins bind individual Tat receptor complexes and are collectively transported. *Embo J* **29**: 1477-1488.

MacCallum, J.L. & D.P. Tieleman, (2011) Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions. *Trends in biochemical sciences* **36**: 653-662.

Maier, T., F. Lottspeich & A. Bock, (1995) GTP hydrolysis by HypB is essential for nickel insertion into hydrogenases of *Escherichia coli*. *Eur J Biochem* **230**: 133-138.

Maillard, J., C.A. Spronk, G. Buchanan, V. Lyall, D.J. Richardson, T. Palmer, G.W. Vuister & F. Sargent, (2007) Structural diversity in twin-arginine signal peptide-binding proteins. *Proc Natl Acad Sci U S A* **104**: 15641-15646.

Mansell, T.J., S.W. Linderman, A.C. Fisher & M.P. DeLisa, (2010) A rapid protein folding assay for the bacterial periplasm. *Protein science : a publication of the Protein Society* **19**: 1079-1090.

Marquis, R.E., K. Mayzel & E.L. Carstensen, (1976) Cation exchange in cell walls of gram-positive bacteria. *Canadian journal of microbiology* **22**: 975-982.

Maurer, C., S. Panahandeh, A.C. Jungkamp, M. Moser & M. Muller, (2010) TatB functions as an oligomeric binding site for folded Tat precursor proteins. *Mol Biol Cell* **21**: 4151-4161.

McCann, J.R., J.A. McDonough, M.S. Pavelka & M. Braunstein, (2007) β-lactamase can function as a reporter of bacterial protein export during *Mycobacterium tuberculosis* infection of host cells. *Microbiol-Sgm* **153**: 3350-3359.

McDonough, J.A., K.E. Hacker, A.R. Flores, M.S. Pavelka, Jr. & M. Braunstein, (2005) The twin-arginine translocation pathway of *Mycobacterium smegmatis* is functional and required for the export of *mycobacterial* β-lactamases. *J Bacteriol* **187**: 7667-7679.

Meindl-Beinker, N.M., C. Lundin, I. Nilsson, S.H. White & G. von Heijne, (2006) Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *Embo Rep* **7**: 1111-1116.

Meloni, S., L. Rey, S. Sidler, J. Imperial, T. Ruiz-Argueso & J.M. Palacios, (2003) The twin-arginine translocation (Tat) system is essential for *Rhizobium*-legume symbiosis. *Mol Microbiol* **48**: 1195-1207.

Mendel, S., A. McCarthy, J.P. Barnett, R.T. Eijlander, A. Nenninger, O.P. Kuipers & C. Robinson, (2008) The *Escherichia coli* TatABC system and a *Bacillus subtilis* TatAC-type system recognise three distinct targeting determinants in twin-arginine signal peptides. *J Mol Biol* **375**: 661-672.

Menetret, J.F., J. Schaletzky, W.M. Clemons, Jr., A.R. Osborne, S.S. Skanland, C. Denison, S.P. Gygi, D.S. Kirkpatrick, E. Park, S.J. Ludtke, T.A. Rapoport & C.W. Akey, (2007) Ribosome binding of a single copy of the SecY complex: implications for protein translocation. *Mol Cell* **28**: 1083-1092.

Mio, K., T. Tsukazaki, H. Mori, M. Kawata, T. Moriya, Y. Sasaki, R. Ishitani, K. Ito, O. Nureki & C. Sato, (2014) Conformational variation of the translocon enhancing chaperone SecDF. *Journal of structural and functional genomics* **15**: 107-115.

Molik, S., I. Karnauchov, C.E. Weidlich, R.G. Herrmann & R.B. Klosgen, (2001) The Rieske Fe/S protein of the cytochrome b_6/*f* complex in chloroplasts - Missing link in the evolution of protein transport pathways in chloroplasts? *Journal of Biological Chemistry* **276**: 42761-42766.

Monne, M., T. Hessa, L. Thissen & G. von Heijne, (2005) Competition between neighboring topogenic signals during membrane protein insertion into the ER. *Febs J* **272**: 28-36.

Morath, S., S. von Aulock & T. Hartung, (2005) Structure/function relationships of lipoteichoic acids. *Journal of endotoxin research* **11**: 348-356.

Mori, H. & K. Cline, (2002) A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid ΔpH/Tat translocase. *J Cell Biol* **157**: 205-210.

Mould, R.M. & C. Robinson, (1991) A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane. *J Biol Chem* **266**: 12189-12193.

Mullineaux, C.W., A. Nenninger, N. Ray & C. Robinson, (2006) Diffusion of green fluorescent protein in three cell environments in *Escherichia coli*. *J Bacteriol* **188**: 3442-3448.

Murakami, H., K. Kita & Y. Anraku, (1986) Purification and properties of a diheme cytochrome b_{561} of the *Escherichia coli* respiratory chain. *J Biol Chem* **261**: 548-551.

Murphy, C.K. & J. Beckwith, (1994) Residues essential for the function of SecE, a membrane component of the *Escherichia coli* secretion apparatus, are located in a conserved cytoplasmic region. *Proc Natl Acad Sci U S A* **91**: 2557-2561.

Musser, S.M. & S.M. Theg, (2000) Characterization of the early steps of OE17 precursor transport by the thylakoid ΔpH/Tat machinery. *Eur J Biochem* **267**: 2588-2598.

Niebisch, A. & M. Bott, (2001) Molecular analysis of the cytochrome bc_1-aa_3 branch of the *Corynebacterium glutamicum* respiratory chain containing an unusual diheme cytochrome c_1. *Arch Microbiol* **175**: 282-294.

Niebisch, A. & M. Bott, (2003) Purification of a cytochrome bc_1-aa_3 supercomplex with quinol oxidase activity from *Corynebacterium glutamicum*. Identification of a fourth subunity of cytochrome aa_3 oxidase and mutational analysis of diheme cytochrome c_1. *J Biol Chem* **278**: 4339-4346.

Nikaido, H., (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and molecular biology reviews : MMBR* **67**: 593-656.

Nilsson, I., A. Saaf, P. Whitley, G. Gafvelin, C. Waller & G. von Heijne, (1998) Proline-induced disruption of a transmembrane α-helix in its natural environment. *J Mol Biol* **284**: 1165-1175.

Nilsson, I. & G. von Heijne, (1990) Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. *Cell* **62**: 1135-1141.

Nilsson, I. & G. von Heijne, (1998) Breaking the camel's back: proline-induced turns in a model transmembrane helix. *J Mol Biol* **284**: 1185-1189.

Nishiyama, K., M. Hanada & H. Tokuda, (1994) Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of *Escherichia coli* at low temperature. *Embo J* **13**: 3272-3277.

Nishiyama, K., S. Mizushima & H. Tokuda, (1993) A novel membrane protein involved in protein translocation across the cytoplasmic membrane of *Escherichia coli*. *Embo J* **12**: 3409-3415.

Nouwen, N., M. van der Laan & A.J. Driessen, (2001) SecDFYajC is not required for the maintenance of the proton motive force. *Febs Lett* **508**: 103-106.

Ojemalm, K., S.C. Botelho, C. Studle & G. von Heijne, (2013) Quantitative analysis of SecYEG-mediated insertion of transmembrane α-helices into the bacterial inner membrane. *J Mol Biol* **425**: 2813-2822.

Ojemalm, K., K.K. Halling, I. Nilsson & G. von Heijne, (2012) Orientational preferences of neighboring helices can drive ER insertion of a marginally hydrophobic transmembrane helix. *Mol Cell* **45**: 529-540.

Oku, Y., K. Kurokawa, M. Matsuo, S. Yamada, B.L. Lee & K. Sekimizu, (2009) Pleiotropic roles of polyglycerolphosphate synthase of lipoteichoic acid in growth of *Staphylococcus aureus* cells. *J Bacteriol* **191**: 141-151.

Oresnik, I.J., C.L. Ladner & R.J. Turner, (2001) Identification of a twin-arginine leader-binding protein. *Mol Microbiol* **40**: 323-331.

Ota, K., M. Sakaguchi, N. Hamasaki & K. Mihara, (2000) Membrane integration of the second transmembrane segment of band 3 requires a closely apposed preceding signal-anchor sequence. *J Biol Chem* **275**: 29743-29748.

Palmer, T. & B.C. Berks, (2012) The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol* **10**: 483-496.

Palmer, T., F. Sargent & B.C. Berks, (2010) The Tat Protein Export Pathway. *EcoSal Plus* **4**.

Panahandeh, S., C. Maurer, M. Moser, M.P. DeLisa & M. Muller, (2008) Following the path of a twin-arginine precursor along the TatABC translocase of *Escherichia coli*. *J Biol Chem* **283**: 33267-33275.

Papanikou, E., S. Karamanou, C. Baud, M. Frank, G. Sianidis, D. Keramisanou, C.G. Kalodimos, A. Kuhn & A. Economou, (2005) Identification of the preprotein binding domain of SecA. *J Biol Chem* **280**: 43209-43217.

Papish, A.L., C.L. Ladner & R.J. Turner, (2003) The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the *Escherichia coli* twin-arginine translocase. *J Biol Chem* **278**: 32501-32506.

Park, E., J.F. Menetret, J.C. Gumbart, S.J. Ludtke, W. Li, A. Whynot, T.A. Rapoport & C.W. Akey, (2014) Structure of the SecY channel during initiation of protein translocation. *Nature* **506**: 102-106.

Park, E. & T.A. Rapoport, (2011) Preserving the membrane barrier for small molecules during bacterial protein translocation. *Nature* **473**: 239-242.

Phillips, G.J. & T.J. Silhavy, (1992) The *E. coli ffh* gene is necessary for viability and efficient protein export. *Nature* **359**: 744-746.

Pickering, B.S. & I.J. Oresnik, (2010) The twin arginine transport system appears to be essential for viability in *Sinorhizobium meliloti*. *J Bacteriol* **192**: 5173-5180.

Plath, K., W. Mothes, B.M. Wilkinson, C.J. Stirling & T.A. Rapoport, (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94**: 795-807.

Pogliano, J.A. & J. Beckwith, (1994) SecD and SecF facilitate protein export in *Escherichia coli*. *Embo J* **13**: 554-561.

Pohlschroder, M., W.A. Prinz, E. Hartmann & J. Beckwith, (1997) Protein translocation in the three domains of life: variations on a theme. *Cell* **91**: 563-566.

Pommier, J., V. Mejean, G. Giordano & C. Iobbi-Nivol, (1998) TorD, a cytoplasmic chaperone that interacts with the unfolded trimethylamine N-oxide reductase enzyme (TorA) in *Escherichia coli*. *J Biol Chem* **273**: 16615-16620.

Porcelli, I., E. de Leeuw, R. Wallis, E. van den Brink-van der Laan, B. de Kruijff, B.A. Wallace, T. Palmer & B.C. Berks, (2002) Characterization and membrane assembly of the TatA component of the *Escherichia coli* twin-arginine protein transport system. *Biochemistry* **41**: 13690-13697.

Pradel, N., J. Delmas, L.F. Wu, C.L. Santini & R. Bonnet, (2009) Sec- and Tat-dependent translocation of β-lactamases across the *Escherichia coli* inner membrane. *Antimicrob Agents Chemother* **53**: 242-248.

Pradel, N., C. Ye, V. Livrelli, J. Xu, B. Joly & L.F. Wu, (2003) Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic *Escherichia coli* O157:H7. *Infection and immunity* **71**: 4908-4916.

Raetz, C.R. & C. Whitfield, (2002) Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**: 635-700.

Ramasamy, S., R. Abrol, C.J. Suloway & W.M. Clemons, Jr., (2013) The glove-like structure of the conserved membrane protein TatC provides insight into signal sequence recognition in twin-arginine translocation. *Structure* **21**: 777-788.

Rao, F., Q. Ji, I. Soehano & Z.X. Liang, (2011) Unusual heme-binding PAS domain from YybT family proteins. *J Bacteriol* **193**: 1543-1551.

Rapoport, T.A., (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**: 663-669.

Rapp, M., E. Granseth, S. Seppala & G. von Heijne, (2006) Identification and evolution of dual-topology membrane proteins. *Nature structural & molecular biology* **13**: 112-116.

Rapp, M., S. Seppala, E. Granseth & G. von Heijne, (2007) Emulating membrane protein evolution by rational design. *Science* **315**: 1282-1284.

Ray, N., J. Oates, R.J. Turner & C. Robinson, (2003) DmsD is required for the biogenesis of DMSO reductase in *Escherichia coli* but not for the interaction of the DmsA signal peptide with the Tat apparatus. *Febs Lett* **534**: 156-160.

Raynaud, C., C. Guilhot, J. Rauzier, Y. Bordat, V. Pelicic, R. Manganelli, I. Smith, B. Gicquel & M. Jackson, (2002) Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. *Mol Microbiol* **45**: 203-217.

Richter, S., U. Lindenstrauss, C. Lucke, R. Bayliss & T. Bruser, (2007) Functional Tat transport of unstructured, small, hydrophilic proteins. *J Biol Chem* **282**: 33257-33264.

Robinson, C., C.F. Matos, D. Beck, C. Ren, J. Lawrence, N. Vasisht & S. Mendel, (2011) Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochim Biophys Acta* **1808**: 876-884.

Rodrigue, A., A. Chanal, K. Beck, M. Muller & L.F. Wu, (1999) Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. *J Biol Chem* **274**: 13223-13228.

Rodriguez, F., S.L. Rouse, C.E. Tait, J. Harmer, A. De Riso, C.R. Timmel, M.S. Sansom, B.C. Berks & J.R. Schnell, (2013) Structural model for the protein-translocating element of the twin-arginine transport system. *Proc Natl Acad Sci U S A* **110**: E1092-1101.

Rollauer, S.E., M.J. Tarry, J.E. Graham, M. Jaaskelainen, F. Jager, S. Johnson, M. Krehenbrink, S.M. Liu, M.J. Lukey, J. Marcoux, M.A. McDowell, F. Rodriguez, P. Roversi, P.J. Stansfeld, C.V. Robinson, M.S.P. Sansom, T. Palmer, M. Hogbom, B.C. Berks & S.M. Lea, (2012) Structure of the TatC core of the twin-arginine protein transport system. *Nature* **492**: 210-+.

Rose, R.W., T. Bruser, J.C. Kissinger & M. Pohlschroder, (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol Microbiol* **45**: 943-950.

Rossier, O. & N.P. Cianciotto, (2005) The *Legionella pneumophila tatB* gene facilitates secretion of phospholipase C, growth under iron-limiting conditions, and intracellular infection. *Infection and immunity* **73**: 2020-2032.

Saaf, A., M. Monne, J.W. de Gier & G. von Heijne, (1998) Membrane topology of the 60-kDa Oxa1p homologue from *Escherichia coli*. *J Biol Chem* **273**: 30415-30418.

Sachelaru, I., N.A. Petriman, R. Kudva, P. Kuhn, T. Welte, B. Knapp, F. Drepper, B. Warscheid & H.G. Koch, (2013) YidC occupies the lateral gate of the SecYEG

translocon and is sequentially displaced by a nascent membrane protein. *J Biol Chem* **288**: 16295-16307.

Saint-Joanis, B., C. Demangel, M. Jackson, P. Brodin, L. Marsollier, H. Boshoff & S.T. Cole, (2006) Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of *Mycobacterium tuberculosis*, increases β-lactam susceptibility and virulence. *J Bacteriol* **188**: 6669-6679.

Samuelson, J.C., M. Chen, F. Jiang, I. Moller, M. Wiedmann, A. Kuhn, G.J. Phillips & R.E. Dalbey, (2000) YidC mediates membrane protein insertion in bacteria. *Nature* **406**: 637-641.

Saparov, S.M., K. Erlandson, K. Cannon, J. Schaletzky, S. Schulman, T.A. Rapoport & P. Pohl, (2007) Determining the conductance of the SecY protein translocation channel for small molecules. *Mol Cell* **26**: 501-509.

Sargent, F., B.C. Berks & T. Palmer, (2002) Assembly of membrane-bound respiratory complexes by the Tat protein-transport system. *Arch Microbiol* **178**: 77-84.

Sargent, F., B.C. Berks & T. Palmer, (2006) Pathfinders and trailblazers: a prokaryotic targeting system for transport of folded proteins. *FEMS microbiology letters* **254**: 198-207.

Sargent, F., E.G. Bogsch, N.R. Stanley, M. Wexler, C. Robinson, B.C. Berks & T. Palmer, (1998) Overlapping functions of components of a bacterial Sec-independent protein export pathway. *Embo J* **17**: 3640-3650.

Sargent, F., U. Gohlke, E. De Leeuw, N.R. Stanley, T. Palmer, H.R. Saibil & B.C. Berks, (2001) Purified components of the *Escherichia coli* Tat protein transport system form a double-layered ring structure. *Eur J Biochem* **268**: 3361-3367.

Sasaki, S., S. Matsuyama & S. Mizushima, (1990) *In vitro* kinetic analysis of the role of the positive charge at the amino-terminal region of signal peptides in translocation of secretory protein across the cytoplasmic membrane in *Escherichia coli*. *J Biol Chem* **265**: 4358-4363.

Sauve, V., S. Bruno, B.C. Berks & A.M. Hemmings, (2007) The SoxYZ complex carries sulfur cycle intermediates on a peptide swinging arm. *J Biol Chem* **282**: 23194-23204.

Schaerlaekens, K., L. Van Mellaert, E. Lammertyn, N. Geukens & J. Anne, (2004) The importance of the Tat-dependent protein secretion pathway in *Streptomyces* as revealed by phenotypic changes in *tat* deletion mutants and genome analysis. *Microbiology* **150**: 21-31.

Schneider, C.A., W.S. Rasband & K.W. Eliceiri, (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**: 671-675.

Schubert, T., O. Lenz, E. Krause, R. Volkmer & B. Friedrich, (2007) Chaperones specific for the membrane-bound [NiFe]-hydrogenase interact with the Tat signal peptide of the small subunit precursor in *Ralstonia eutropha* H16. *Mol Microbiol* **66**: 453-467.

Schulze, R.J., J. Komar, M. Botte, W.J. Allen, S. Whitehouse, V.A. Gold, A.N.J.A. Lycklama, K. Huard, I. Berger, C. Schaffitzel & I. Collinson, (2014) Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proc Natl Acad Sci U S A* **111**: 4844-4849.

Scott, J.R. & T.C. Barnett, (2006) Surface proteins of Gram-positive bacteria and how they get there. *Annu Rev Microbiol* **60**: 397-423.

Scotti, P.A., M.L. Urbanus, J. Brunner, J.W. de Gier, G. von Heijne, C. van der Does, A.J. Driessen, B. Oudega & J. Luirink, (2000) YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. *Embo J* **19**: 542-549.

Seiler, P., T. Ulrichs, S. Bandermann, L. Pradl, S. Jorg, V. Krenn, L. Morawietz, S.H. Kaufmann & P. Aichele, (2003) Cell-wall alterations as an attribute of *Mycobacterium tuberculosis* in latent infection. *The Journal of infectious diseases* **188**: 1326-1331.

Seppala, S., J.S. Slusky, P. Lloris-Garcera, M. Rapp & G. von Heijne, (2010) Control of membrane protein topology by a single C-terminal residue. *Science* **328**: 1698-1700.

Settles, A.M., A. Yonetani, A. Baron, D.R. Bush, K. Cline & R. Martienssen, (1997) Sec-independent protein translocation by the maize Hcf106 protein. *Science* **278**: 1467-1470.

Shan, S.O., S. Chandrasekar & P. Walter, (2007) Conformational changes in the GTPase modules of the signal reception particle and its receptor drive initiation of protein translocation. *J Cell Biol* **178**: 611-620.

Shanmugham, A., H.W. Wong Fong Sang, Y.J. Bollen & H. Lill, (2006) Membrane binding of twin arginine preproteins as an early step in translocation. *Biochemistry* **45**: 2243-2249.

Shaw, A.S., P.J. Rottier & J.K. Rose, (1988) Evidence for the loop model of signal-sequence insertion into the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **85**: 7592-7596.

Sianidis, G., S. Karamanou, E. Vrontou, K. Boulias, K. Repanas, N. Kyrpides, A.S. Politou & A. Economou, (2001) Cross-talk between catalytic and regulatory elements in a DEAD motor domain is essential for SecA function. *Embo J* **20**: 961-970.

Silhavy, T.J., D. Kahne & S. Walker, (2010) The bacterial cell envelope. *Cold Spring Harbor perspectives in biology* **2**: a000414.

Sone, N., K. Nagata, H. Kojima, J. Tajima, Y. Kodera, T. Kanamaru, S. Noguchi & J. Sakamoto, (2001) A novel hydrophobic diheme *c*-type cytochrome. Purification from *Corynebacterium glutamicum* and analysis of the *Qcr*CBA operon encoding three subunit proteins of a putative cytochrome reductase complex. *Biochim Biophys Acta* **1503**: 279-290.

Stanley, N.R., K. Findlay, B.C. Berks & T. Palmer, (2001) *Escherichia coli* strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. *J Bacteriol* **183**: 139-144.

Stanley, N.R., T. Palmer & B.C. Berks, (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. *Journal of Biological Chemistry* **275**: 11591-11596.

Stanley, N.R., F. Sargent, G. Buchanan, J. Shi, V. Stewart, T. Palmer & B.C. Berks, (2002) Behaviour of topological marker proteins targeted to the Tat protein transport pathway. *Mol Microbiol* **43**: 1005-1021.

Strauch, E.M. & G. Georgiou, (2007) *Escherichia coli tatC* mutations that suppress defective twin-arginine transporter signal peptides. *J Mol Biol* **374**: 283-291.

Tam, P.C., A.P. Maillard, K.K. Chan & F. Duong, (2005) Investigating the SecY plug movement at the SecYEG translocation channel. *Embo J* **24**: 3380-3388.

Tarry, M.J., E. Schafer, S. Chen, G. Buchanan, N.P. Greene, S.M. Lea, T. Palmer, H.R. Saibil & B.C. Berks, (2009) Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. *P Natl Acad Sci USA* **106**: 13284-13289.

Taubert, J., B. Hou, H.J. Risselada, D. Mehner, H. Lunsdorf, H. Grubmuller & T. Bruser, (2015) TatBC-independent TatA/Tat substrate interactions contribute to transport efficiency. *PLoS One* **10**: e0119761.

Thony-Meyer, L. & P. Kunzler, (1997) Translocation to the periplasm and signal sequence cleavage of preapocytochrome *c* depend on *sec* and *lep*, but not on the *ccm* gene products. *Eur J Biochem* **246**: 794-799.

Tottey, S., K.J. Waldron, S.J. Firbank, B. Reale, C. Bessant, K. Sato, T.R. Cheek, J. Gray, M.J. Banfield, C. Dennison & N.J. Robinson, (2008) Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* **455**: 1138-1142.

Trun, N.J., J. Stader, A. Lupas, C. Kumamoto & T.J. Silhavy, (1988) Two cellular components, PrlA and SecB, that recognize different sequence determinants are required for efficient protein export. *J Bacteriol* **170**: 5928-5930.

Tsubaki, M., F. Takeuchi & N. Nakanishi, (2005) Cytochrome b_{561} protein family: expanding roles and versatile transmembrane electron transfer abilities as predicted by a new classification system and protein sequence motif analyses. *Biochim Biophys Acta* **1753**: 174-190.

Tsukazaki, T., H. Mori, Y. Echizen, R. Ishitani, S. Fukai, T. Tanaka, A. Perederina, D.G. Vassylyev, T. Kohno, A.D. Maturana, K. Ito & O. Nureki, (2011) Structure and function of a membrane component SecDF that enhances protein export. *Nature* **474**: 235-238.

Tullman-Ercek, D., M.P. DeLisa, Y. Kawarasaki, P. Iranpour, B. Ribnicky, T. Palmer & G. Georgiou, (2007) Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides. *Journal of Biological Chemistry* **282**: 8309-8316.

Uehara, T., K.R. Parzych, T. Dinh & T.G. Bernhardt, (2010) Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *Embo J* **29**: 1412-1422.

Urbanus, M.L., P.A. Scotti, L. Froderberg, A. Saaf, J.W. de Gier, J. Brunner, J.C. Samuelson, R.E. Dalbey, B. Oudega & J. Luirink, (2001) Sec-dependent membrane protein insertion: sequential interaction of nascent FtsQ with SecY and YidC. *Embo Rep* **2**: 524-529.

Valent, Q.A., D.A. Kendall, S. High, R. Kusters, B. Oudega & J. Luirink, (1995) Early events in preprotein recognition in *E. coli*: interaction of SRP and trigger factor with nascent polypeptides. *Embo J* **14**: 5494-5505.

van Alphen, L., A. Verkleij, J. Leunissen-Bijvelt & B. Lugtenberg, (1978) Architecture of the outer membrane of *Escherichia coli*. III. Protein-lipopolysaccharide complexes in intramembraneous particles. *J Bacteriol* **134**: 1089-1098.

Van den Berg, B., W.M. Clemons, Jr., I. Collinson, Y. Modis, E. Hartmann, S.C. Harrison & T.A. Rapoport, (2004) X-ray structure of a protein-conducting channel. *Nature* **427**: 36-44.

van der Laan, M., N.P. Nouwen & A.J. Driessen, (2005) YidC-an evolutionary conserved device for the assembly of energy-transducing membrane protein complexes. *Current opinion in microbiology* **8**: 182-187.

Vasil, M.L., A.P. Tomaras & A.E. Pritchard, (2012) Identification and evaluation of twin-arginine translocase inhibitors. *Antimicrob Agents Chemother* **56**: 6223-6234.

Verelst, W. & H. Asard, (2003) A phylogenetic study of cytochrome b_{561} proteins. *Genome biology* **4**: R38.

Viklund, H., E. Granseth & A. Elofsson, (2006) Structural classification and prediction of reentrant regions in α-helical transmembrane proteins: application to complete genomes. *J Mol Biol* **361**: 591-603.

Virkki, M.T., C. Peters, D. Nilsson, T. Sorensen, S. Cristobal, B. Wallner & A. Elofsson, (2014) The positive inside rule is stronger when followed by a transmembrane helix. *J Mol Biol* **426**: 2982-2991.

Vlasuk, G.P., S. Inouye, H. Ito, K. Itakura & M. Inouye, (1983) Effects of the complete removal of basic amino acid residues from the signal peptide on secretion of lipoprotein in *Escherichia coli*. *J Biol Chem* **258**: 7141-7148.

Vollmer, W., D. Blanot & M.A. de Pedro, (2008) Peptidoglycan structure and architecture. *FEMS microbiology reviews* **32**: 149-167.

von Heijne, G., (1989) Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**: 456-458.

von Heijne, G., (1990) The signal peptide. *The Journal of membrane biology* **115**: 195-201.

von Heijne, G., (1997) Getting greasy: How transmembrane polypeptide segments integrate into the lipid bilayer. *Mol Microbiol* **24**: 249-253.

von Heijne, G., (2006) Membrane-protein topology. *Nat Rev Mol Cell Bio* **7**: 909-918.

von Heijne, G. & Y. Gavel, (1988) Topogenic signals in integral membrane proteins. *Eur J Biochem* **174**: 671-678.

Voorhees, R.M., I.S. Fernandez, S.H. Scheres & R.S. Hegde, (2014) Structure of the mammalian ribosome-Sec61 complex to 3.4 A resolution. *Cell* **157**: 1632-1643.

Voorhees, R.M. & R.S. Hegde, (2016) Structure of the Sec61 channel opened by a signal sequence. *Science* **351**: 88-91.

Voulhoux, R., G. Ball, B. Ize, M.L. Vasil, A. Lazdunski, L.F. Wu & A. Filloux, (2001) Involvement of the twin-arginine translocation system in protein secretion *via* the type II pathway. *Embo J* **20**: 6735-6741.

Wahlberg, J.M. & M. Spiess, (1997) Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain. *J Cell Biol* **137**: 555-562.

Waterhouse, A.M., J.B. Procter, D.M. Martin, M. Clamp & G.J. Barton, (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

Weiner, J.H., P.T. Bilous, G.M. Shaw, S.P. Lubitz, L. Frost, G.H. Thomas, J.A. Cole & R.J. Turner, (1998) A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell* **93**: 93-101.

Wexler, M., F. Sargent, R.L. Jack, N.R. Stanley, E.G. Bogsch, C. Robinson, B.C. Berks & T. Palmer, (2000) TatD is a cytoplasmic protein with DNase activity - No requirement for TatD family proteins in Sec-independent protein export. *Journal of Biological Chemistry* **275**: 16717-16722.

White, S.H. & G. von Heijne, (2008) How translocons select transmembrane helices. *Annu Rev Biophys* **37**: 23-42.

Widdick, D.A., K. Dilks, G. Chandra, A. Bottrill, M. Naldrett, M. Pohlschroder & T. Palmer, (2006) The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*. *P Natl Acad Sci USA* **103**: 17927-17932.

Widdick, D.A., R.T. Eijlander, J.M. van Dijl, O.P. Kuipers & T. Palmer, (2008) A facile reporter system for the experimental identification of twin-arginine translocation (Tat) signal peptides from all kingdoms of life. *J Mol Biol* **375**: 595-603.

Xia, D., C.A. Yu, H. Kim, J.Z. Xia, A.M. Kachurin, L. Zhang, L. Yu & J. Deisenhofer, (1997) Crystal structure of the cytochrome bc_1 complex from bovine heart mitochondria. *Science* **277**: 60-66.

Xie, K., T. Hessa, S. Seppala, M. Rapp, G. von Heijne & R.E. Dalbey, (2007) Features of transmembrane segments that promote the lateral release from the translocase into the lipid phase. *Biochemistry* **46**: 15153-15161.

Xu, Z., J.D. Knafels & K. Yoshino, (2000) Crystal structure of the bacterial protein export chaperone secB. *Nature structural biology* **7**: 1172-1177.

Yahr, T.L. & W.T. Wickner, (2001) Functional reconstitution of bacterial Tat translocation *in vitro*. *The EMBO journal* **20**: 2472-2479.

Yamagishi, M., Y. Onishi, S. Yoshimura, H. Fujita, K. Imai, Y. Kida & M. Sakaguchi, (2014) A Few Positively Charged Residues Slow Movement of a Polypeptide Chain across the Endoplasmic Reticulum Membrane. *Biochemistry* **53**: 5375-5383.

Yan, C. & J. Luo, (2010) An analysis of reentrant loops. *The protein journal* **29**: 350-354.

Yau, W.M., W.C. Wimley, K. Gawrisch & S.H. White, (1998) The preference of tryptophan for membrane interfaces. *Biochemistry* **37**: 14713-14718.

Yen, M.R., Y.H. Tseng, E.H. Nguyen, L.F. Wu & M.H. Saier, Jr., (2002) Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. *Arch Microbiol* **177**: 441-450.

Yu, Z., G. Koningstein, A. Pop & J. Luirink, (2008) The conserved third transmembrane segment of YidC contacts nascent *Escherichia coli* inner membrane proteins. *J Biol Chem* **283**: 34635-34642.

Zhang, L., Y. Sato, T. Hessa, G. von Heijne, J.K. Lee, I. Kodama, M. Sakaguchi & N. Uozumi, (2007) Contribution of hydrophobic and electrostatic interactions to the

membrane integration of the Shaker K+ channel voltage sensor domain. *Proc Natl Acad Sci U S A* **104**: 8263-8268.

Zhang, Y., L. Wang, Y. Hu & C. Jin, (2014) Solution structure of the TatB component of the twin-arginine translocation system. *Biochim Biophys Acta* **1838**: 1881-1888.

Zhang, Z., L. Huang, V.M. Shulmeister, Y.I. Chi, K.K. Kim, L.W. Hung, A.R. Crofts, E.A. Berry & S.H. Kim, (1998) Electron transfer by domain movement in cytochrome bc_1. *Nature* **392**: 677-684.

Zhu, L., C. Klenner, A. Kuhn & R.E. Dalbey, (2012) Both YidC and SecYEG are required for translocation of the periplasmic loops 1 and 2 of the multispanning membrane protein TatC. *J Mol Biol* **424**: 354-367.

Zhu, L., A. Wasey, S.H. White & R.E. Dalbey, (2013) Charge composition features of model single-span membrane proteins that determine selection of YidC and SecYEG translocase pathways in *Escherichia coli*. *J Biol Chem* **288**: 7704-7716.

Zimmer, J., Y. Nam & T.A. Rapoport, (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**: 936-943.

Zoufaly, S., J. Frobel, P. Rose, T. Flecken, C. Maurer, M. Moser & M. Muller, (2012) Mapping precursor-binding site on TatC subunit of twin arginine-specific protein translocase by site-specific photo cross-linking. *J Biol Chem* **287**: 13430-13441.

8 Appendix

8.1 Oligonucleotides used in this study

Table 8.1 Oligonucleotides used in this study for generation of constructs.
All oligonucleotides unless designated are forward (white) reverse (grey).

Primer	Sequence	Restriction sites	Use
***Bam*HI Bla**	GCGCGGATCCATGCACCCAGAAACGCTGG	*Bam*HI	For construction of vector containing Bla
***Bam*HI AmiA**	GCGCGGATCCATGAGCACTTTTAAACCACTAAAA	*Bam*HI	For construction of vector containing AmiA
SU18.1	CGTATGTTGTGTGGAATTGTGAGC		For sequencing of vector pSU18 & pSUPROM
SU18.2	GGGTAACGCCAGGGTTTTCCC		For sequencing of vector pSU18 & pSUPROM
Uni-Rep1-Hind	GCGCAAGCTTTGTCGGTTGGCGCAAAACACGCTG		For sequencing of Tat promoter
Inside Bla reverse	GCTGCAGGCATCGTGGTGTCACGCTCGTC		For sequencing within Bla
Inside AmiA reverse	TCAAACAGCACTTGTTGCAATAGGTGATC		For sequencing within AmiA
***Mycobacterium tuberculosis* Rieske**	GCGCGGATCCATGAGCCGCGCGGATGAT	*Bam*HI	For construction of vector containing Mycobacterium tuberculosis extended Rieske
***Mycobacterium tuberculosis* Rieske extension 243aa**	GCGCTCTAGACGCCTTCTTGCCTTCCGCGGTCGGCACCACCGGTTTCCACGGGTTTTTAATCAGGCCGCCCGCAAA	*Xba*I	For construction of vector containing extended MTB Rieske
Rieske forward	GCGCGGATCCATGAGTAGCCAAGACATT	*Bam*HI	For construction of vector containing extended Rieske
Rieske extension 205aa reverse	GCGCTCTAGAGAGCTTGCCCTTGGACCA	*Xba*I	For construction of vector containing extended Rieske
Rieske extension 205aa	GCGCTCTAGAGAGCTTGCCCTTGGACCACAGGGTGTGGCGGAGCTTGGTCCCGGGCAGCGGACCGAGGTCGCGCAGCAGGACGACGCC	*Xba*I	For construction of vector containing extended Rieske
Rieske extension 204aa	GCGCTCTAGACTTGCCCTTGGACCACAGGGTGTGGCGGAGCTTGGTCCCGGGCAGCGGACCGAGGTCGCGCAGCAGGACGACGCC	*Xba*I	For construction of vector containing extended Rieske

Rieske extension 206aa	GCGCTCTAGAGATGAGCTTGCCCTTGGACCACAGGGTGTGGCGGA GCTTGGTCCCGGGCAGCGGACCGAGGTCGCGCAGCAGGACGACG CC	*Xba*I	For construction of vector containing extended Rieske
Rieske extension 198aa	GCGCTCTAGAGGTGTGGCGGAGCTTGGTCCCGGGCAGCGGACCGA GGTCGCGCAGCAGGACGACGCC	*Xba*I	For construction of vector containing extended Rieske
Rieske RHHR	GTCGCCGACGAGCATCGCCCGATCGAGGCG		For Rieske Quikchange
	CGCCTCGATCGGGCGATGCTCGTCGGCGAC		
Rieske RHKK	GTCGCCGACGAGAAAAAACCGATCGAGGCG		For Rieske Quikchange
	CGCCTCGATCGGTTTTTTCTCGTCGGCGAC		
Rieske A138P	CGTCACCCGATCGAGCCGTCCCCCGAGGTCCGT		For Rieske Quikchange
	ACGGACCTCGGGGGACGGCTCGATCGGGTGACG		
Rieske A144P	TCCCCCGAGGTCCGTCCCAAGGTCCACGCGGAC		For Rieske Quikchange
	GTCCGCGTGGACCTTGGGACGGACCTCGGGGGA		
Rieske A148P	CGTGCCAAGGTCCACCCGGACTTCAAGCAGGGT		For Rieske Quikchange
	ACCCTGCTTGAAGTCCGGGTGGACCTTGGCACG		
Rieske A154P	GACTTCAAGCAGGGTCCCAAGGAGTCCGTGATC		For Rieske Quikchange
	CAGCACGGACTCCTTGGGACCCTGCTTGAAGTC		
Rieske M124A	TGGGCCCGCACCCTGGCCTCCGACGAGGAGGTC		For Rieske Quikchange
	GACCTCCTCGTCGGAGGCCAGGGTGCGGGCCCA		
Rieske S125A	GCCCGCACCCTGATGGCCGACGAGGAGGTCGCC		For Rieske Quikchange
	GGCGACCTCCTCGTCGGCCATCAGGGTGCGGGC		
Rieske D126A	CGCACCCTGATGTCCGCCGAGGAGGTCGCCGAC		For Rieske Quikchange
	GTCGGCGACCTCCTCGGCGGACATCAGGGTGCG		
Rieske E127A	ACCCTGATGTCCGACGCCGAGGTCGCCGACGAG		For Rieske Quikchange
	CTCGTCGGCGACCTCGGCGTCGGACATCAGGGT		
Rieske M124L	TGGGCCCGCACCCTGTTGTCCGACGAGGAGGTC		For Rieske Quikchange
	GACCTCCTCGTCGGACAACAGGGTGCGGGCCCA		
Rieske S125L	GCCCGCACCCTGATGTTGGACGAGGAGGTCGCC		For Rieske Quikchange
	GGCGACCTCCTCGTCCAACATCAGGGTGCGGGC		
Rieske D126L	CGCACCCTGATGTCCTTGGAGGAGGTCGCCGAC		For Rieske Quikchange
	GTCGGCGACCTCCTCCAAGGACATCAGGGTGCG		

Rieske E127L	ACCCTGATGTCCGACTTGGAGGTCGCCGACGAG	For Rieske Quikchange
	CTCGTCGGCGACCTCCAAGTCGGACATCAGGGT	
Rieske RRKK	GCCAAGGAGTCCGTGATCGGGAAGAAGAAGCTGATCCGCAACACG	For Rieske Quikchange
	CGTGTTGCGGATCAGCTTCTTCTTCCCGATCACGGACTCCTTGGC	
Rieske RRKQ	GCCAAGGAGTCCGTGATCGGGAAGCAGAAGCTGATCCGCAACACG	For Rieske Quikchange
	CGTGTTGCGGATCAGCTTCTGCTTCCCGATCACGGACTCCTTGGC	
Rieske RRAA	GCCAAGGAGTCCGTGATCGGGGCCGCCAAGCTGATCCGCAACACG	For Rieske Quikchange
	CGTGTTGCGGATCAGCTTGGCGGCCCCGATCACGGACTCCTTGGC	
Rieske RRAD	GCCAAGGAGTCCGTGATCGGGGCCGATAAGCTGATCCGCAACACG	For Rieske Quikchange
	CGTGTTGCGGATCAGCTTATCGGCCCCGATCACGGACTCCTTGGC	
Rieske ∆RR	TCCGTGATCGGGAAGCTGATCCGCAACACGATGCTG	For Rieske Quikchange
	CGCCTAGTCGAAGGGCTAGTGCCTGAGGAACCGTGG	
Rieske ∆118-122	GCGGGCGCGGTCCTGATGTCCGACGAGGAGGTCGCC	For Rieske Modified Quikchange
	GTCGGACATCAGGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆123-127	TGGGCCCGCACCGAGGTCGCCGACGAGCGTCACCCG	For Rieske Modified Quikchange
	GTCGGCGACCTCGGTGCGGGCCCAGTGGACCGCGCC	
Rieske ∆128-132	ATGTCCGACGAGCGTCACCCGATCGAGGCGTCCCCC	For Rieske Modified Quikchange
	GATCGGGTGACGCTCGTCGGACATCAGGGTGCGGGC	
Rieske ∆133-137	GTCGCCGACGAGGCGTCCCCCGAGGTCCGTGCCAAG	For Rieske Modified Quikchange
	CTCGGGGGACGCGTCGGCGACCTCCTCGTCGGACAT	
Rieske ∆138-142	CACCCGATCGAGCGTGCCAAGGTCCACGCGGACTTC	For Rieske Modified Quikchange
	GACCTTGGCACGCTCGATCGGGTGACGCTCGTCGGC	
Rieske ∆143-147	TCCCCCGAGGTCGCGGACTTCAAGCAGGGTGCCAAG	For Rieske Modified Quikchange
	CTTGAAGTCCGCGACCTCGGGGGACGCCTCGATCGG	
Rieske ∆148-152	GCCAAGGTCCACGGTGCCAAGGAGTCCGTGATCGGG	For Rieske Modified Quikchange
	CTCCTTGGCACCGTGGACCTTGGCACGGACCTCGGG	
Rieske ∆153-157	GACTTCAAGCAGGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACCTGCTTGAAGTCCGCGTGGACCTT	
Rieske ∆157-162	GCCAAGGAGTCCAAGCTGATCCGCAACACGATGCTG	For Rieske Modified Quikchange
	GCGGATCAGCTTGGACTCCTTGGCACCCTGCTTGAA	
Rieske ∆131-132	GAGGAGGTCGCCCGTCACCCGATCGAGGCGTCCCCC	For Rieske Modified Quikchange

	GATCGGGTGACGGGCGACCTCCTCGTCGGACATCAG	
Rieske ∆126-128	ACCCTGATGTCCGTCGCCGACGAGCGTCACCCGATC	For Rieske Modified Quikchange
	CTCGTCGGCGACGGACATCAGGGTGCGGGCCCAGTG	
Rieske ∆137 ∆141	CACCCGATCGCGTCCCCCGTCCGTGCCAAGGTCCACGCG	For Rieske Modified Quikchange
	GGCACGGACGGGGGACGCGATCGGGTGACGCTCGTCGGC	
Rieske ∆149 ∆156	AAGGTCCACGCGTTCAAGCAGGGTGCCAAGTCCGTGATCGGG	For Rieske Modified Quikchange
	CCCGATCACGGACTTGGCACCCTGCTTGAACGCGTGGACCTT	
Rieske E137A E141A	CGTCACCCGATCGCGGCGTCCCCCGCGGTCCGTGCCAAG	For Rieske Quikchange
	CTTGGCACGGACCGCGGGGGACGCCGCGATCGGGTGACG	
Rieske D131A E132A	GACGAGGAGGTCGCCGCGGCGCGTCACCCGATCGAG	For Rieske Quikchange
	CTCGATCGGGTGACGCGCCGCGGCGACCTCCTCGTC	
Rieske D126A E127A E128A	GCCCGCACCCTGATGTCCGCGGCGGCGGTCGCCGACGAGCGTCAC	For Rieske Quikchange
	GTGACGCTCGTCGGCGACCGCCGCCGCGGACATCAGGGTGCGGGC	
Rieske ∆118-127	GCGGGCGCGGTCGAGGTCGCCGACGAGCGTCACCCGATC	For Rieske Modified Quikchange
	GTCGGCGACCTCGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆128-137	ATGTCCGACGAGGCGTCCCCCGAGGTCCGTGCCAAG	For Rieske Modified Quikchange
	CTCGGGGGACGCCTCGTCGGACATCAGGGTGCGGGC	
Rieske ∆138-147	CACCCGATCGAGGCGGACTTCAAGCAGGGTGCCAAG	For Rieske Modified Quikchange
	CTTGAAGTCCGCCTCGATCGGGTGACGCTCGTCGGC	
Rieske ∆148-152	GCCAAGGTCCACGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACGTGGACCTTGGCACGGACCTCGGG	
Rieske Ins D129 E130 E131	TCCGACGAGGAGGACGAGGAGGTCGCCGACGAGCGTCACCCGATC	For Rieske Modified Quikchange
	CTCGTCGGCGACCTCCTCGTCCTCCTCGTCGGACATCAGGGTGCG	
Rieske ∆126-8 ∆131-2	ATGTCCGTCGCCCGTCACCCGATCGAGGCGTCCCCC	For Rieske Modified Quikchange
	GATCGGGTGACGGGCGACGGACATCAGGGTGCGGGC	
Rieske ∆126-7	ACCCTGATGTCCGAGGTCGCCGACGAGCGTCACCCG	For Rieske Modified Quikchange
	GTCGGCGACCTCGGACATCAGGGTGCGGGCCCAGTG	
Rieske ∆127-8	CTGATGTCCGACGTCGCCGACGAGCGTCACCCGATC	For Rieske Modified Quikchange

	CTCGTCGGCGACGTCGGACATCAGGGTGCGGGCCCA	
Rieske-R185C-Bla	TCCGGCGTCGTCCTGCTGTGCTCTAGACACCCAGAAACG	For Rieske Quikchange
	CGTTTCTGGGTGTCTAGAGCACAGCAGGACGACGCCGGA	
Rieske ∆118-132	GCGGGCGCGGTCCGTCACCCGATCGAGGCGTCCCCC	For Rieske Modified Quikchange
	GATCGGGTGACGGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆123-137	TGGGCCCGCACCGCGTCCCCCGAGGTCCGTGCCAAG	For Rieske Modified Quikchange
	CTCGGGGGACGCGGTGCGGGCCCAGTGGACCGCGCC	
Rieske ∆128-142	ATGTCCGACGAGCGTGCCAAGGTCCACGCGGACTTC	For Rieske Modified Quikchange
	GACCTTGGCACGCTCGTCGGACATCAGGGTGCGGGC	
Rieske ∆133-147	GTCGCCGACGAGGCGGACTTCAAGCAGGGTGCCAAG	For Rieske Modified Quikchange
	CTTGAAGTCCGCCTCGTCGGCGACCTCCTCGTCGGA	
Rieske ∆138-152	CACCCGATCGAGGGTGCCAAGGAGTCCGTGATCGGG	For Rieske Modified Quikchange
	CTCCTTGGCACCCTCGATCGGGTGACGCTCGTCGGC	
Rieske ∆143-157	TCCCCCGAGGTCGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACGACCTCGGGGGACGCCTCGATCGG	
Rieske ∆118-137	GCGGGCGCGGTCGCGTCCCCCGAGGTCCGTGCCAAG	For Rieske Modified Quikchange
	CTCGGGGGACGCGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆138-157	CACCCGATCGAGGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACCTCGATCGGGTGACGCTCGTCGGC	
Rieske ∆131-2 with template ∆126-8 ∆137 ∆141	GCGCATGTCCGTCGCCCGTCACCCGATCGCGTCCCCCGTC	For Rieske Modified Quikchange
	GCGCGATCGGGTGACGGGCGACGGACATCAGGGTGCGGGC	
Rieske D126K E127K E128K	GCGCACCCTGATGTCCAAGAAGAAGGTCGCCGACGAGCGTCACCCG	For Rieske Quikchange
	GCGCCTCGTCGGCGACCTTCTTCTTGGACATCAGGGTGCGGGCCCA	
Rieske D131K E132K	GAGGAGGTCGCCAAGAAGCGTCACCCGATCGAGGCGTCC	For Rieske Quikchange
	GATCGGGTGACGCTTCTTGGCGACCTCCTCGTCGGACAT	
Rieske E137K E141K	CGTCACCCGATCAAGGCGTCCCCCAAGGTCCGTGCCAAGGTCCACGCG	For Rieske Quikchange

	CTTGGCACGGACCTTGGGGGACGCCTTGATCGGGTGACGCTCGTC GGC	
Rieske ∆118-142	GCGGGCGCGGTCCGTGCCAAGGTCCACGCGGACTTC	For Rieske Modified Quikchange
	GACCTTGGCACGGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆123-147	TGGGCCCGCACCGCGGACTTCAAGCAGGGTGCCAAG	For Rieske Modified Quikchange
	CTTGAAGTCCGCGGTGCGGGCCCAGTGGACCGCGCC	
Rieske ∆128-152	ATGTCCGACGAGGGTGCCAAGGAGTCCGTGATCGGG	For Rieske Modified Quikchange
	CTCCTTGGCACCCTCGTCGGACATCAGGGTGCGGGC	
Rieske ∆133-157	GTCGCCGACGAGGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACCTCGTCGGCGACCTCCTCGTCGGA	
Rieske ∆118-147	GCGGGCGCGGTCGCGGACTTCAAGCAGGGTGCCAAG	For Rieske Modified Quikchange
	CTTGAAGTCCGCGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆123-152	TGGGCCCGCACCGGTGCCAAGGAGTCCGTGATCGGG	For Rieske Modified Quikchange
	CTCCTTGGCACCGGTGCGGGCCCAGTGGACCGCGCC	
Rieske ∆128-157	ATGTCCGACGAGGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACCTCGTCGGACATCAGGGTGCGGGC	
Rieske D131KE132K with template D126KE127K E128K E137KE141K	AAGAAGGTCGCCAAGAAGCGTCACCCGATCAAGGCGTCC	For Rieske Quikchange
	GATCGGGTGACGCTTCTTGGCGACCTTCTTCTTGGACAT	
Rieske ∆118-152	GCGGGCGCGGTCGGTGCCAAGGAGTCCGTGATCGGG	For Rieske Modified Quikchange
	CTCCTTGGCACCGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆123-157	TGGGCCCGCACCGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACGGTGCGGGCCCAGTGGACCGCGCC	
Rieske ∆118-157	GCGGGCGCGGTCGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACGACCGCGCCCGCGCCGATGGCGAA	
Rieske-R185A-Bla	TCCGGCGTCGTCCTGCTGGCGTCTAGACACCCAGAAACG	For Rieske Quikchange
	CGTTTCTGGGTGTCTAGACGCCAGCAGGACGACGCCGGA	
Rieske ∆118-153	GCGGGCGCGGTCGCCAAGGAGTCCGTGATCGGGCGC	For Rieske Modified Quikchange
	GGACTCCTTGGCGACCGCGCCCGCGCCGATGGCGAA	

Rieske ∆118-154	GCGGGCGCGGTCAAGGAGTCCGTGATCGGGCGCCGC	For Rieske Modified Quikchange
	CACGGACTCCTTGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆118-155	GCGGGCGCGGTCGAGTCCGTGATCGGGCGCCGCAAG	For Rieske Modified Quikchange
	GATCACGGACTCGACCGCGCCCGCGCCGATGGCGAA	
Rieske ∆118-156	GCGGGCGCGGTCTCCGTGATCGGGCGCCGCAAGCTG	For Rieske Modified Quikchange
	CCCGATCACGGAGACCGCGCCCGCGCCGATGGCGAA	
Rieske S179LG180L	ACCCTGGTGCCGCTCCTGCTGGTCGTCCTGCTGCGC	For Rieske Quikchange
	GCGCAGCAGGACGACCAGCAGGAGCGGCACCAGGGT	
Rieske P177L S179L G180L	GCGCTCACCCTGGTGCTGCTCCTGCTGGTCGTCCTGCTGCGC	For Rieske Quikchange
	GCGCAGCAGGACGACCAGCAGGAGCAGCACCAGGGTGAGCGC	
Rieske V158C	GGTGCCAAGGAGTCCTGCATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATGCAGGACTCCTTGGCACC	
Rieske ∆118-153 V158C	GTCGCCAAGGAGTCCTGCATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATGCAGGACTCCTTGGCGAC	
Rieske ∆118-154 V158C	GCGGTCAAGGAGTCCTGCATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATGCAGGACTCCTTGACCGC	
Rieske ∆118-155 V158C	GGCGCGGTCGAGTCCTGCATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATGCAGGACTCGACCGCGCC	
Rieske ∆118-156 V158C	GCGGGCGCGGTCTCCTGCATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATGCAGGAGACCGCGCCCGC	
Rieske ∆118-157 V158C	GGCGCGGGCGCGGTCTGCATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATGCAGACCGCGCCCGCGCC	
Rieske V158K	GGTGCCAAGGAGTCCAAGATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATCTTGGACTCCTTGGCACC	

Rieske ∆118-153 V158K	GTCGCCAAGGAGTCCAAGATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATCTTGGACTCCTTGGCGAC	
Rieske ∆118-154 V158K	GCGGTCAAGGAGTCCAAGATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATCTTGGACTCCTTGACCGC	
Rieske ∆118-155 V158K	GGCGCGGTCGAGTCCAAGATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATCTTGGACTCGACCGCGCC	
Rieske ∆118-156 V158K	GCGGGCGCGGTCTCCAAGATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATCTTGGAGACCGCGCCCGC	
Rieske ∆118-157 V158K	GGCGCGGGCGCGGTCAAGATCGGGCGCCGCAAG	For Rieske Quikchange
	CTTGCGGCGCCCGATCTTGACCGCGCCCGCGCC	
Rieske V82C	TTCGTGGCGATCGACTGTGACAAGTCGGTCTAC	For Rieske Quikchange
	GTAGACCGACTTGTCACAGTCGATCGCCACGAA	
Rieske S179L	ACCCTGGTGCCGCTCCTGGGCGTCGTCCTGCTGCGC	For Rieske Quikchange
	GCGCAGCAGGACGACGCCCAGGAGCGGCACCAGGGT	
Rieske G180L	ACCCTGGTGCCGCTCTCCCTGGTCGTCCTGCTGCGC	For Rieske Quikchange
	GCGCAGCAGGACGACCAGGGAGAGCGGCACCAGGGT	
Rieske Δ124-157	GCCCGCACCCTGGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACCAGGGTGCGGGCCCAGTGGACCGC	
Rieske Δ125-157	CGCACCCTGATGGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACCATCAGGGTGCGGGCCCAGTGGAC	
Rieske Δ126-157	ACCCTGATGTCCGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACGGACATCAGGGTGCGGGCCCAGTG	
Rieske Δ127-157	CTGATGTCCGACGTGATCGGGCGCCGCAAGCTGATC	For Rieske Modified Quikchange
	GCGCCCGATCACGTCGGACATCAGGGTGCGGGCCCA	
Rieske Loop HWKES	GCGGTCCACTGGAAGGAGTCCGTGATCGGGCGCCGC	For Rieske Modified Quikchange

	CACGGACTCCTTCCAGTGGACCGCGCCCGCGCCGAT	
Rieske Loop HWAKES	GCGGTCCACTGGGCCAAGGAGTCCGTGATCGGGCGCCGC	For Rieske Modified Quikchange
	CACGGACTCCTTGGCCCAGTGGACCGCGCCCGCGCCGAT	
Rieske Loop HWES	GCGGTCCACTGGGAGTCCGTGATCGGGCGCCGC	For Rieske Modified Quikchange
	CACGGACTCCCAGTGGACCGCGCCCGCGCCGAT	
Mycobacterium tuberculosis Rieske RRAD	GGCAGCACCATTCGCGCCGACAAACTGATTGGCCTG	For MTB Rieske Quikchange
	CAGGCCAATCAGTTTGTCGGCGCGAATGGTGCTGCC	

8.2 Rieske-Bla quantification Western blots

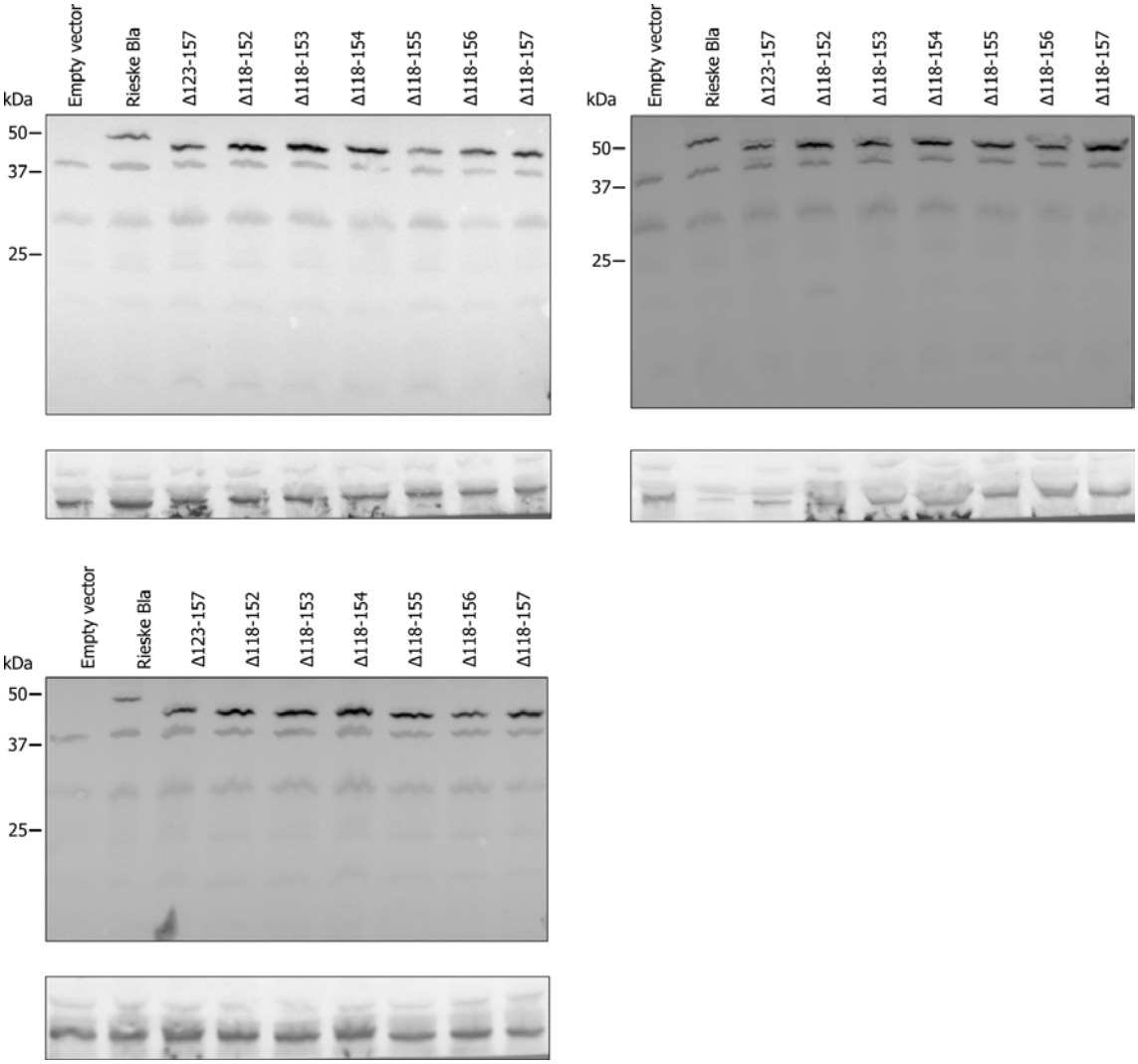


Figure 8.1 Protein is still stably produced from all truncation variants.
In triplicate, biological replicates of crude membrane extracts of DADE (*tat*) harbouring the pSUPROM vector alone (Empty vector) or the pSUPROM vector encoding Rieske-Bla or loop truncation variants (Δ123-157-Rieske-Bla, Δ118-152-Rieske-Bla, Δ118-153-Rieske-Bla, Δ118-154-Rieske-Bla, Δ118-155-Rieske-Bla, Δ118-156-Rieske-Bla and Δ118-157-Rieske-Bla) were separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Rieske peptide antibody (top panels). As a loading control the levels of the OM protein BamA were detected using an anti-BamA antiserum (bottom panels).

8.3 Rieske-AmiA quantification Western blots

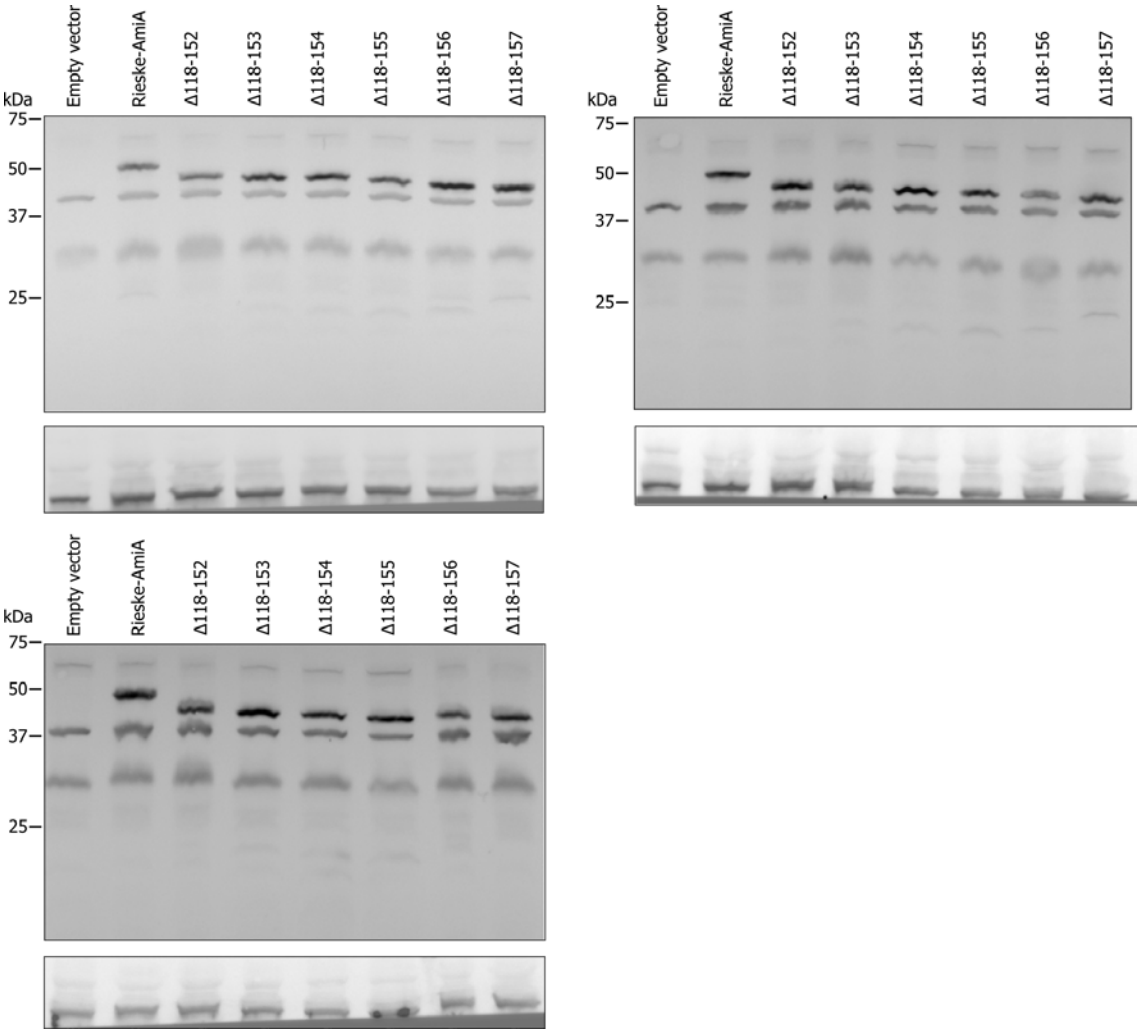


Figure 8.2 Protein is still stably produced from all truncation variants.
In triplicate, biological replicates of crude membrane extracts of MCDSSAC (*tat*$^+$) harbouring the pSUPROM vector alone (Empty vector) or the pSUPROM vector encoding Rieske-AmiA or loop truncation variants Δ118-152-Rieske- AmiA, Δ118-153-Rieske-AmiA, Δ118-154-Rieske-AmiA, Δ118-155-Rieske-AmiA, Δ118-156-Rieske-AmiA and Δ118-157-Rieske-AmiA were separated on SDS-PAGE (12% acrylamide), electroblotted and immunoreactive bands were detected with anti-Rieske peptide antibody (top panels). As a loading control the levels of the OM protein BamA were detected using an anti-BamA antiserum (bottom panels).

8.4 Rieske-AmiA SDS tests

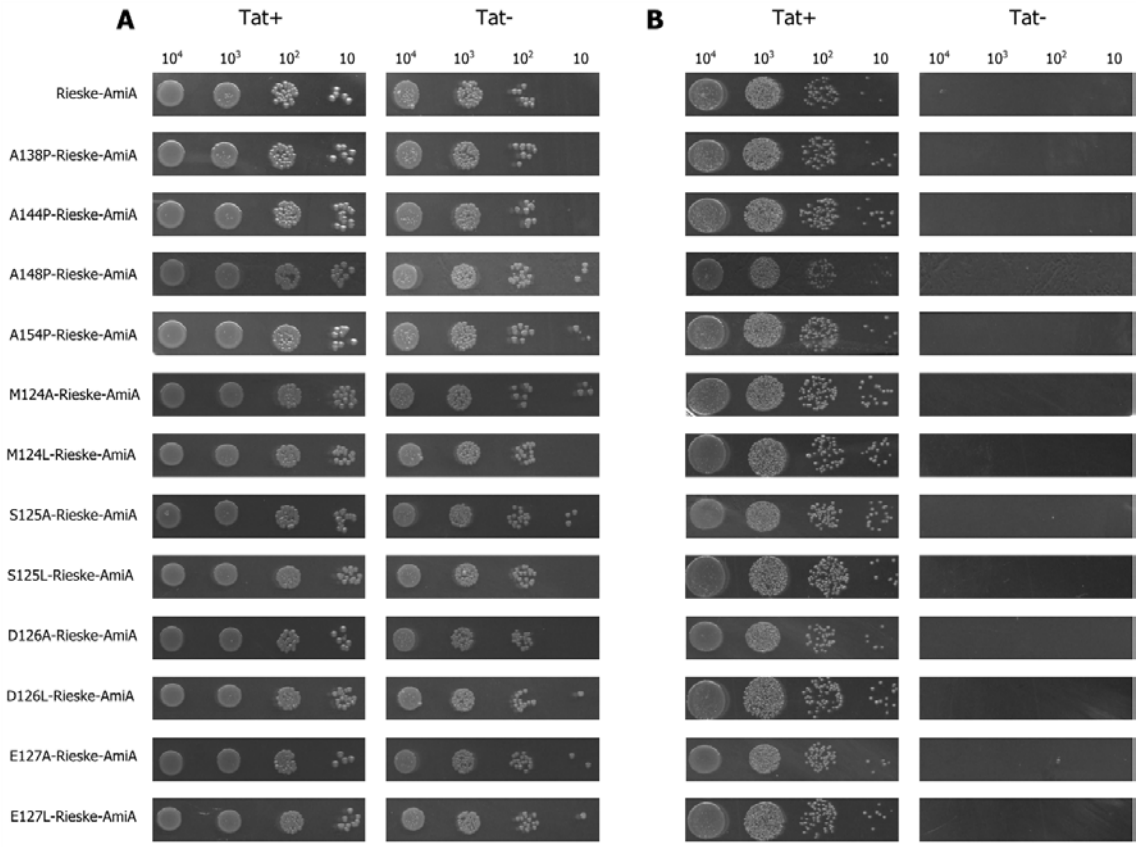


Figure 8.3 SDS test analysis of predicted α-helix variants and negative charge region variants. Representative spot tests of strains MCDSSAC (*tat*$^{+}$ strain) and MCDSSACΔtat (*tat*$^{-}$ strain) harbouring the pSUPROM vector encoding Rieske-AmiA or predicited α–helix fusion variants (A138P-Rieske-AmiA, A144P-Rieske-AmiA, A148P-Rieske-AmiA and A154P-Rieske-AmiA) or negative charge region fusion variants (M124A-Rieske-AmiA, M124L-Rieske-AmiA, S125A-Rieske-AmiA, S125A-Rieske-AmiA, D126A-Rieske-AmiA, D126A-Rieske-AmiA E127A-Rieske-AmiA and E127L-Rieske-AmiA). Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto LB solid agar (A) and LB solid agar containing 1% (w/v) SDS (B). LB agar plates were incubated at 37°C for 16 h.

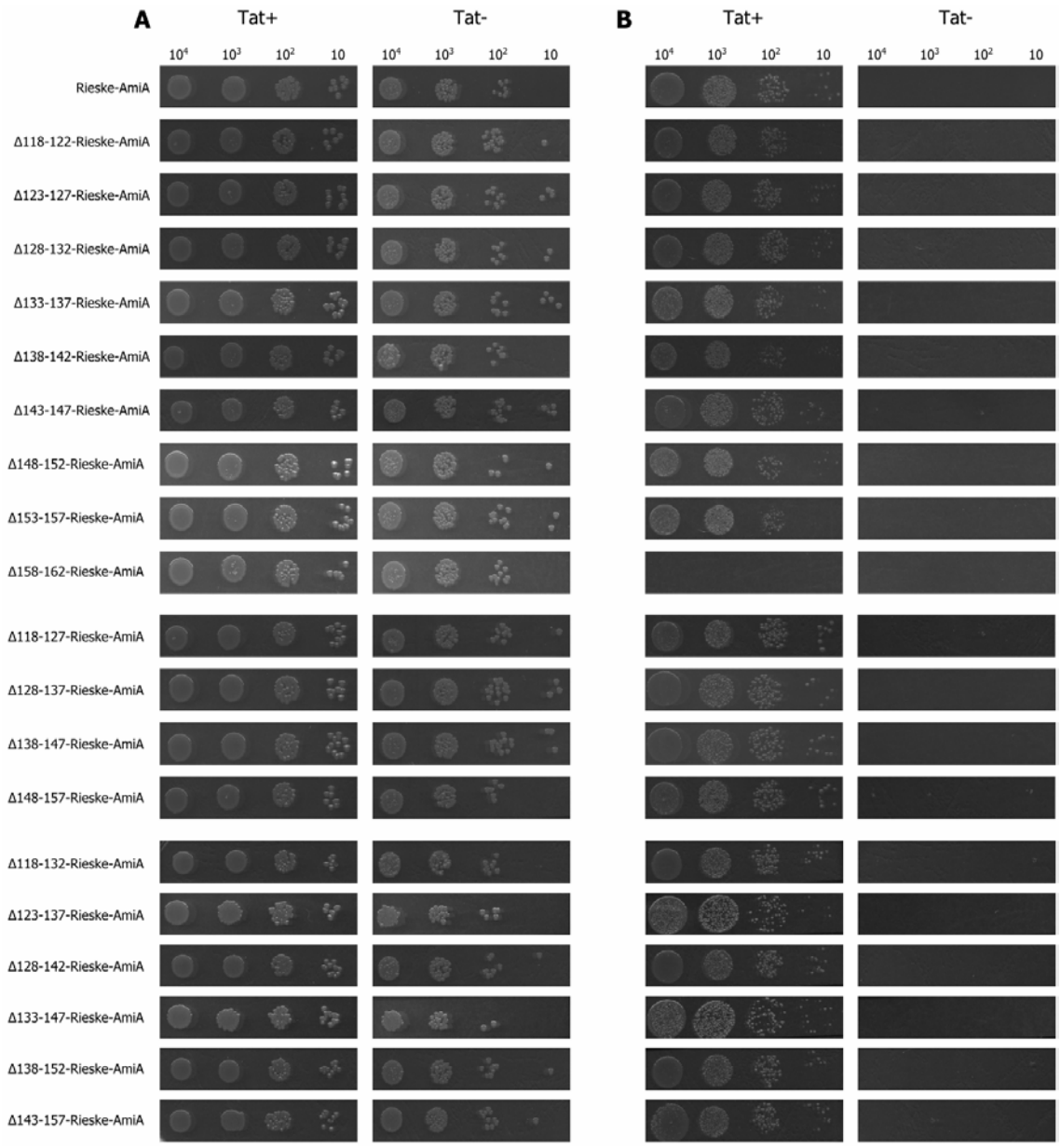


Figure 8.4 SDS test analysis of truncation variants of 5, 10 and 15 residues.
Representative spot tests of strains MCDSSAC (*tat*$^{+}$ strain) and MCDSSACΔtat (*tat*$^{-}$ strain) harbouring the pSUPROM vector encoding Rieske-AmiA or 5 aa deletion fusion variants (Δ118-122-Rieske-AmiA, Δ123-127-Rieske-AmiA, Δ128-132-Rieske-AmiA, Δ133-137-Rieske-AmiA, Δ138-142-Rieske-AmiA, Δ143-147-Rieske-AmiA, Δ148-152-Rieske-AmiA Δ153-157-Rieske-AmiA and Δ158-162-Rieske-AmiA) or 10 aa deletion fusion variants (Δ118-127-Rieske-AmiA, Δ128-137-Rieske-AmiA, Δ138-147-Rieske-AmiA, and Δ148-157-Rieske-AmiA) or 15 aa deletion fusion variants (Δ118-132-Rieske-AmiA, Δ123-137-Rieske-AmiA, Δ128-142-Rieske-AmiA, Δ133-147-Rieske-AmiA, Δ138-152-Rieske-AmiA, and Δ143-157-Rieske-AmiA). Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto LB solid agar (A) and LB solid agar containing 1% (w/v) SDS (B). LB agar plates were incubated at 37°C for 16 h.

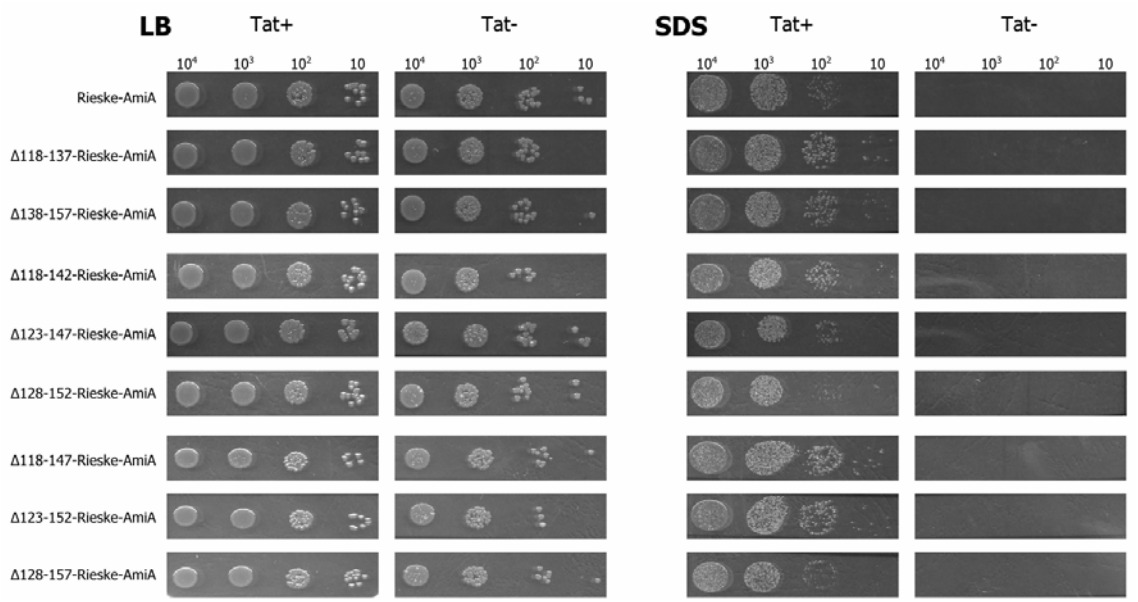
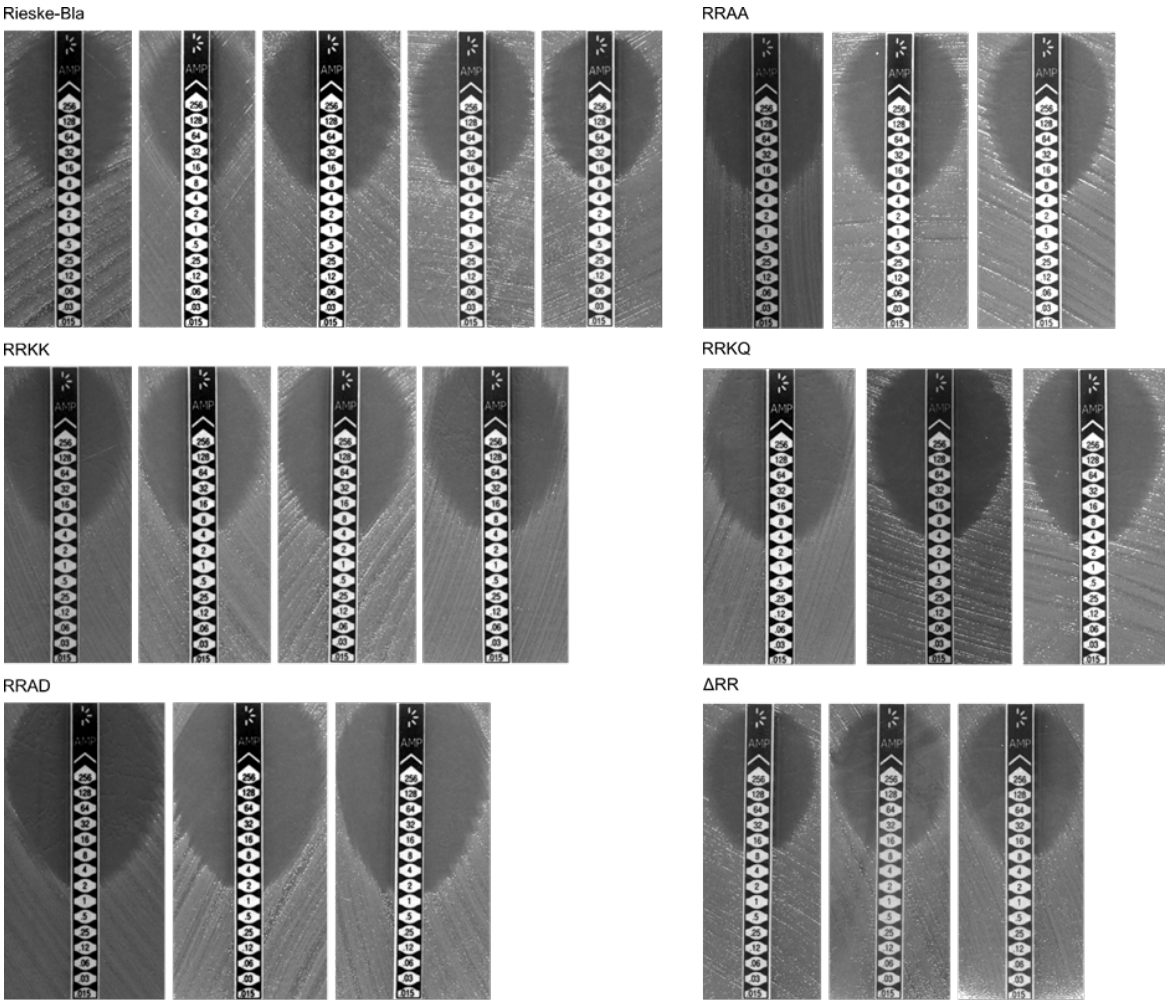


Figure 8.5 SDS test analysis of truncation variants of 20, 25 and 30 residues.
Representative spot tests of strains MCDSSAC (*tat*$^{+}$ strain) and MCDSSACΔtat (*tat*$^{-}$ strain) harbouring the pSUPROM vector encoding Rieske-AmiA or 20 aa deletion fusion variants (Δ118-137-Rieske-AmiA and Δ138-157-Rieske-AmiA) or 25 aa deletion fusion variants (Δ118-142-Rieske-AmiA, Δ123-147-Rieske-AmiA, and Δ128-152-Rieske-AmiA) or 30 aa deletion fusion variants (Δ118-147-Rieske-AmiA, Δ123-152-Rieske-AmiA, and Δ128-157-Rieske-AmiA). Strains were grown overnight in liquid media, diluted to give serial dilutions of 10, 10^2, 10^3 and 10^4 cells per 5 µl aliquots which were spotted onto LB solid agar (A) and LB solid agar containing 1% (w/v) SDS (B). LB agar plates were incubated at 37°C for 16 h.

8.5 Rieske-Bla MIC tests

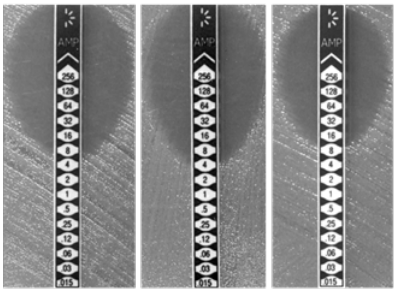


Variant	Mean M.I.C.
Rieske-Bla	7.6 (0.9, n=5)
RRKK	3.8 (0.5, n=4)
RRAD	1.3 (0.3, n=3)
RRAA	5.5 (1.0, n=4)
RRKQ	3.3 (0.6, n=3)
ΔRR	9.3 (2.3, n=3)

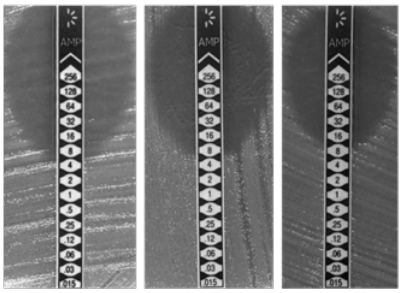
Figure 8.6 M.I.C. analysis of Twin-arginine variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding Rieske-Bla and twin-arginine variants (RRKK-Rieske-Bla, RRAD-Rieske-Bla, RRAA-Rieske-Bla, RRKQ-Rieske-Bla and ΔRR-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The table indicates the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

Negatively charged

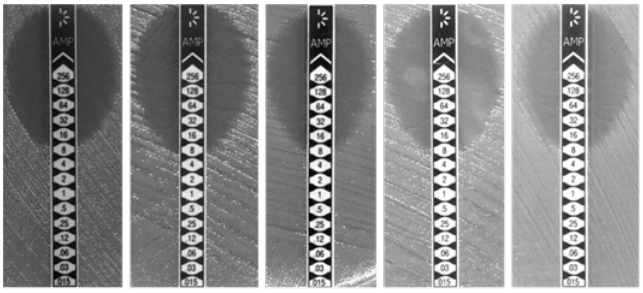
M124A



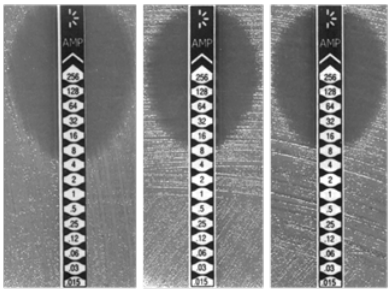
M124L



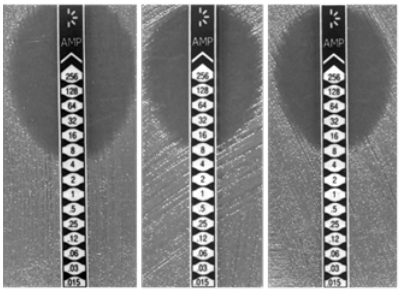
S125A



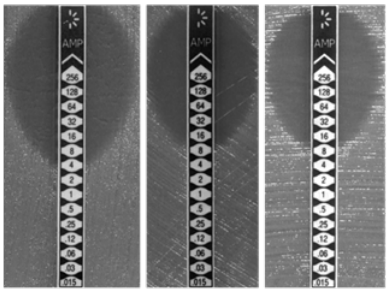
S125L



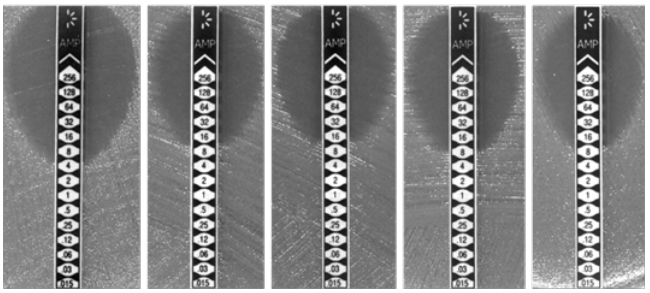
D126A



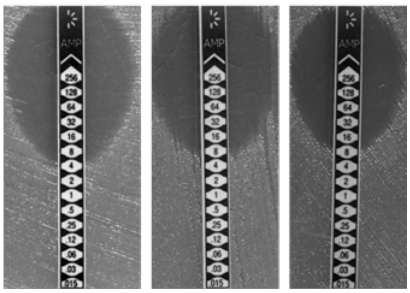
D126L



E127A



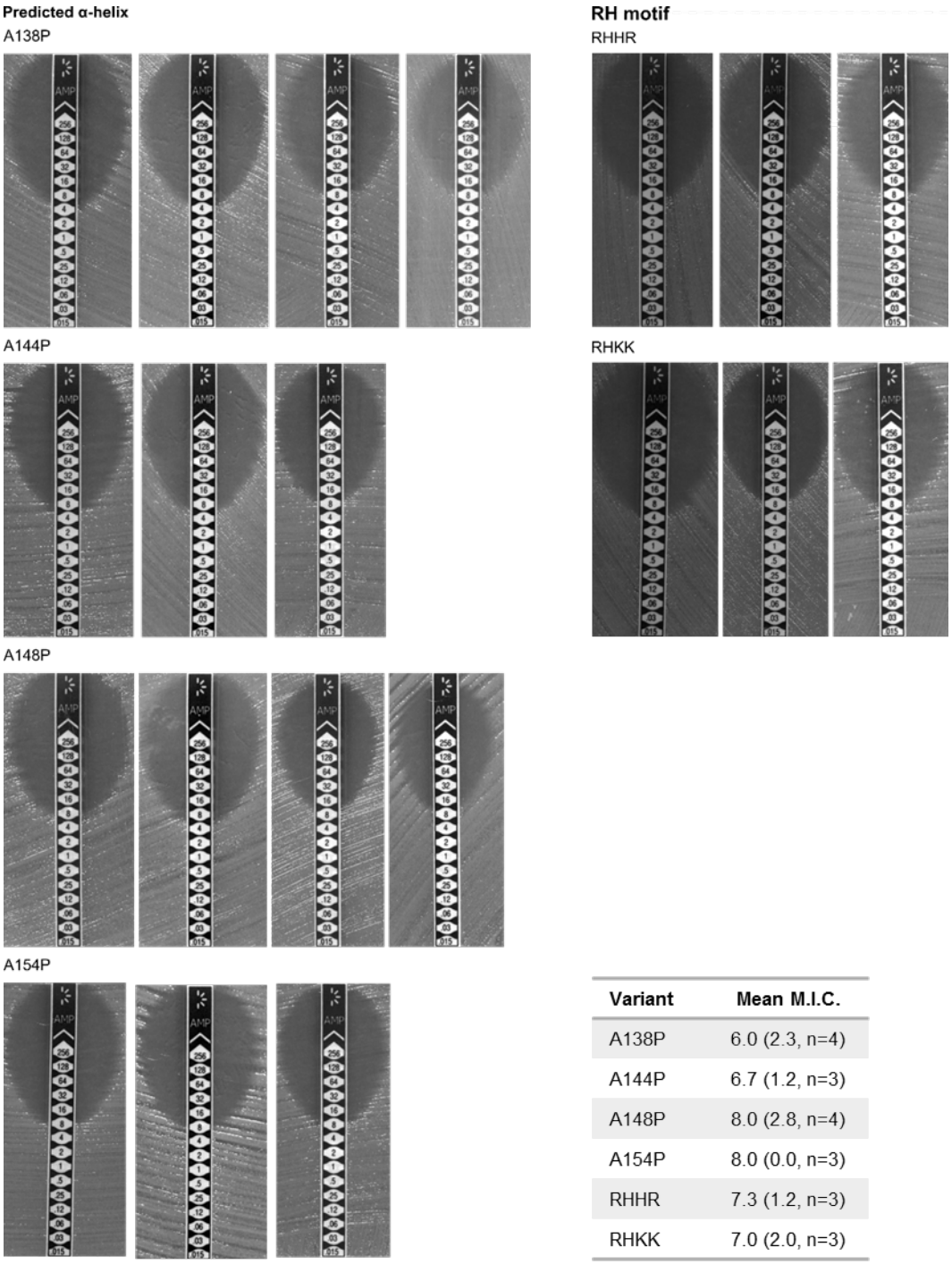
E127L



Variant	Mean M.I.C.
M124A	6.0 (2.0, n=3)
S125A	8.0 (0.0, n=5)
D125A	7.3 (1.2, n=3)
E127A	8.0 (2.8, n=5)

Variant	Mean M.I.C.
M124L	6.0 (2.0, n=3)
S125L	6.7 (2.3, n=3)
D125L	7.3 (1.2, n=3)
E127L	6.0 (2.0, n=3)

Figure 8.7 M.I.C. analysis of negatively charged region variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding negatively charged region variants (M124A-Rieske-Bla, S125A-Rieske-Bla, D126A-Rieske-Bla, E127A-Rieske-Bla M124L-Rieske-Bla, S125L-Rieske-Bla, D126L-Rieske-Bla and E127L-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

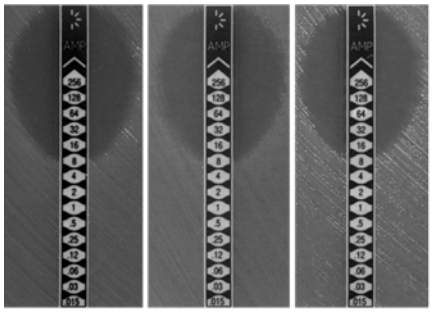


Variant	Mean M.I.C.
A138P	6.0 (2.3, n=4)
A144P	6.7 (1.2, n=3)
A148P	8.0 (2.8, n=4)
A154P	8.0 (0.0, n=3)
RHHR	7.3 (1.2, n=3)
RHKK	7.0 (2.0, n=3)

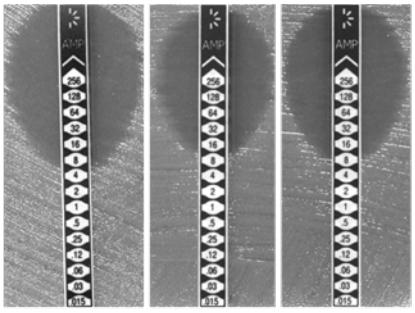
Figure 8.8 M.I.C. analysis of predicted α-helix and RH motif variants.
M.I.C.Evaluator™ strip test of DADE (*tat*$^-$) harbouring the pSUPROM vector encoding predicted α-helix (A138P-Rieske-Bla, A144P-Rieske-Bla, A148P-Rieske-Bla, A154P-Rieske-Bla) or RH variants (RHHR-Rieske-Bla, and RHKK-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (μg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

Removal of negative charges

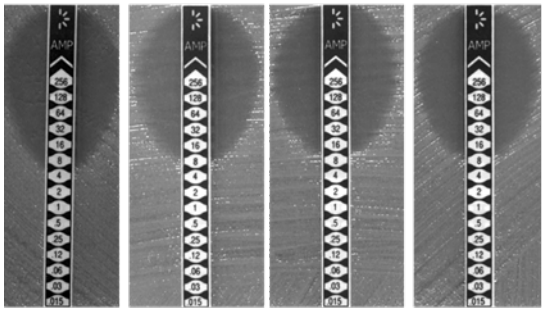
Δ126-128



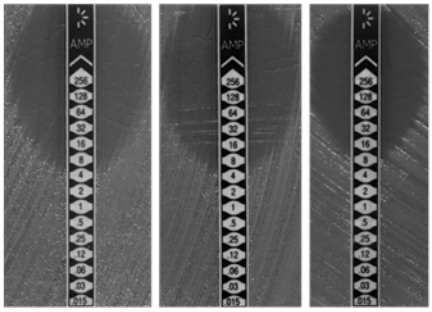
Δ131-132



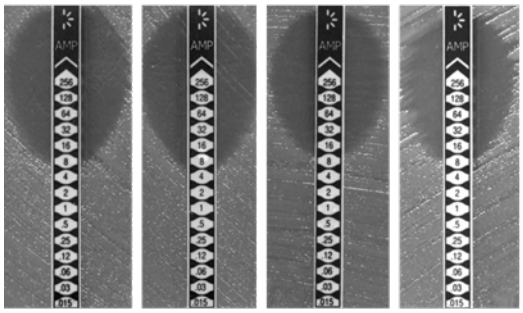
Δ126-127



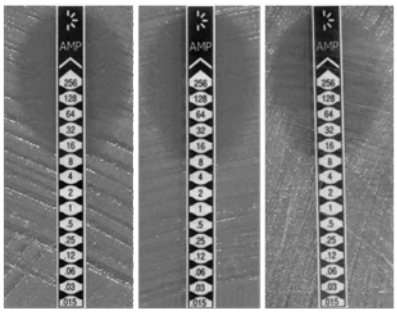
Δ137 Δ141



Δ127-128



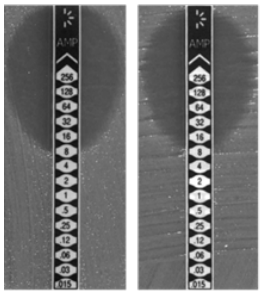
Δ149 Δ156



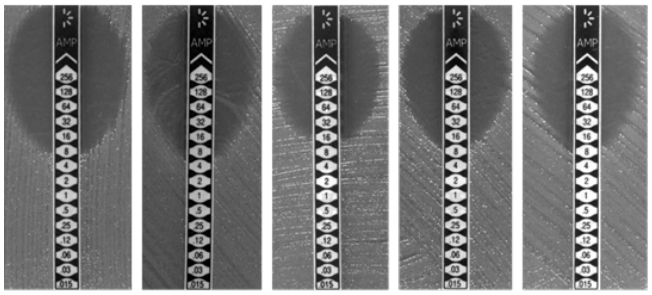
Variant	Mean M.I.C.
Δ126-128	8.0 (0.0, n=3)
Δ126-127	7.5 (1.0, n=4)
Δ127-128	8.0 (2.8, n=4)
Δ131-132	7.3 (1.2, n=3)
Δ137 Δ141	4.0 (0.0, n=3)
Δ149 Δ156	8.0 (0.0, n=3)

Figure 8.9 M.I.C. analysis of truncations of negatively charged variants.
M.I.C.Evaluator™ strip test of DADE (*tat*⁻) harbouring the pSUPROM vector encoding deletions of negatively charged variants (Δ126-128-Rieske-Bla, Δ126-127-Rieske-Bla, Δ127-128-Rieske-Bla, Δ131-132-Rieske-Bla, Δ137Δ141-Rieske-Bla and Δ149Δ156-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

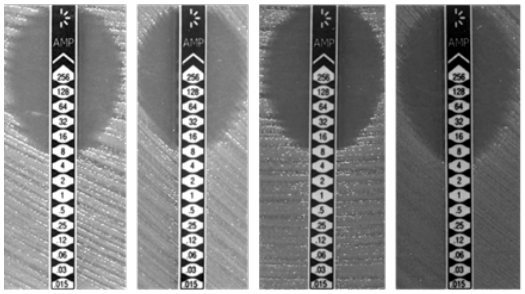
Δ126-128 Δ137 Δ141



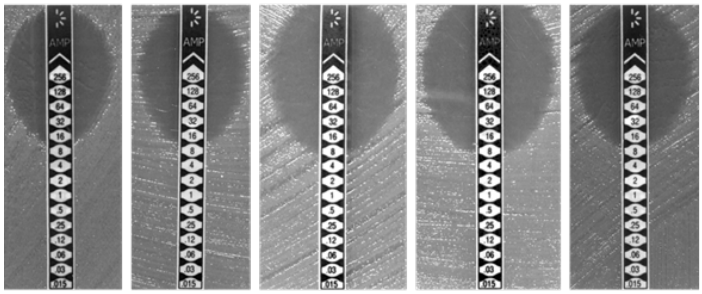
Δ126-128 Δ131-132



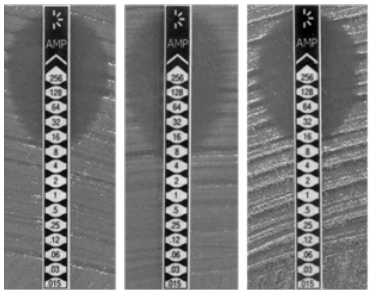
Δ131-132 Δ137 Δ141



Δ126-128 Δ131-132 Δ137 Δ141



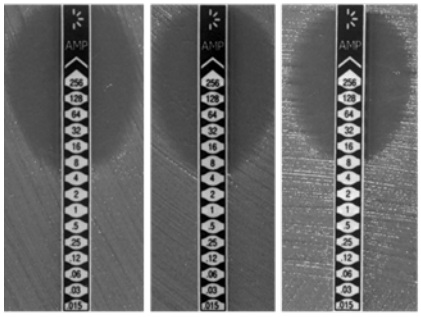
Δ126-128 Δ131-132 Δ137 Δ141 Δ149 Δ156



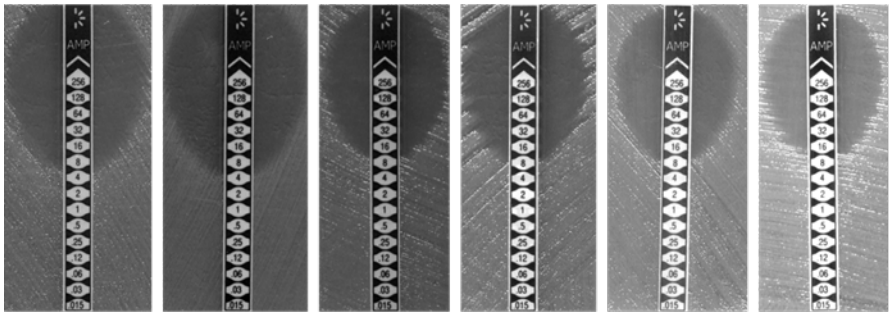
Variant	Mean M.I.C.
Δ126-128 Δ137Δ141	8.0 (0.0, n=2)
Δ126-128 Δ131-132	8.0 (2.4, n=5)
Δ131-132 Δ137Δ141	6.0 (2.0, n=4)
Δ126-127 Δ131-132 Δ137 Δ141	10.4 (2.2, n=5)
Δ126-127 Δ131-132 Δ137 Δ141 Δ149 Δ156	10.0 (2.8, n=3)

Figure 8.10 M.I.C. analysis of truncations of negatively charged variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding deletions of negatively charged variants (Δ128-128Δ137Δ141-Rieske-Bla, Δ126-128Δ131-132-Rieske-Bla, Δ131-132Δ137Δ141-Rieske-Bla, Δ126-128Δ131-132Δ137Δ141-Rieske-Bla and Δ126-128Δ131-132Δ137Δ141Δ149Δ156-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (μg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The table indicates the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

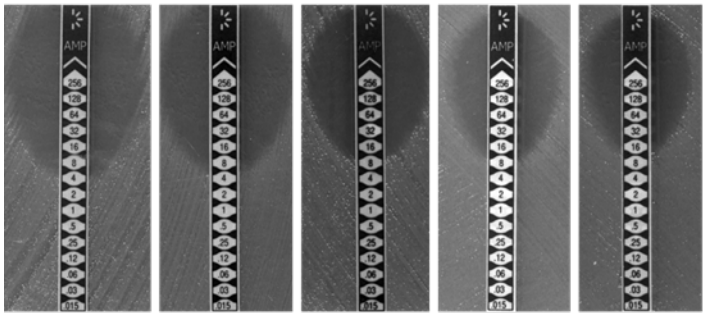
Ins D129 E130 E131



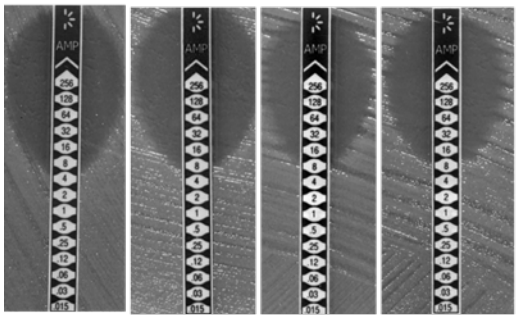
D131A E132A



E137A E141A

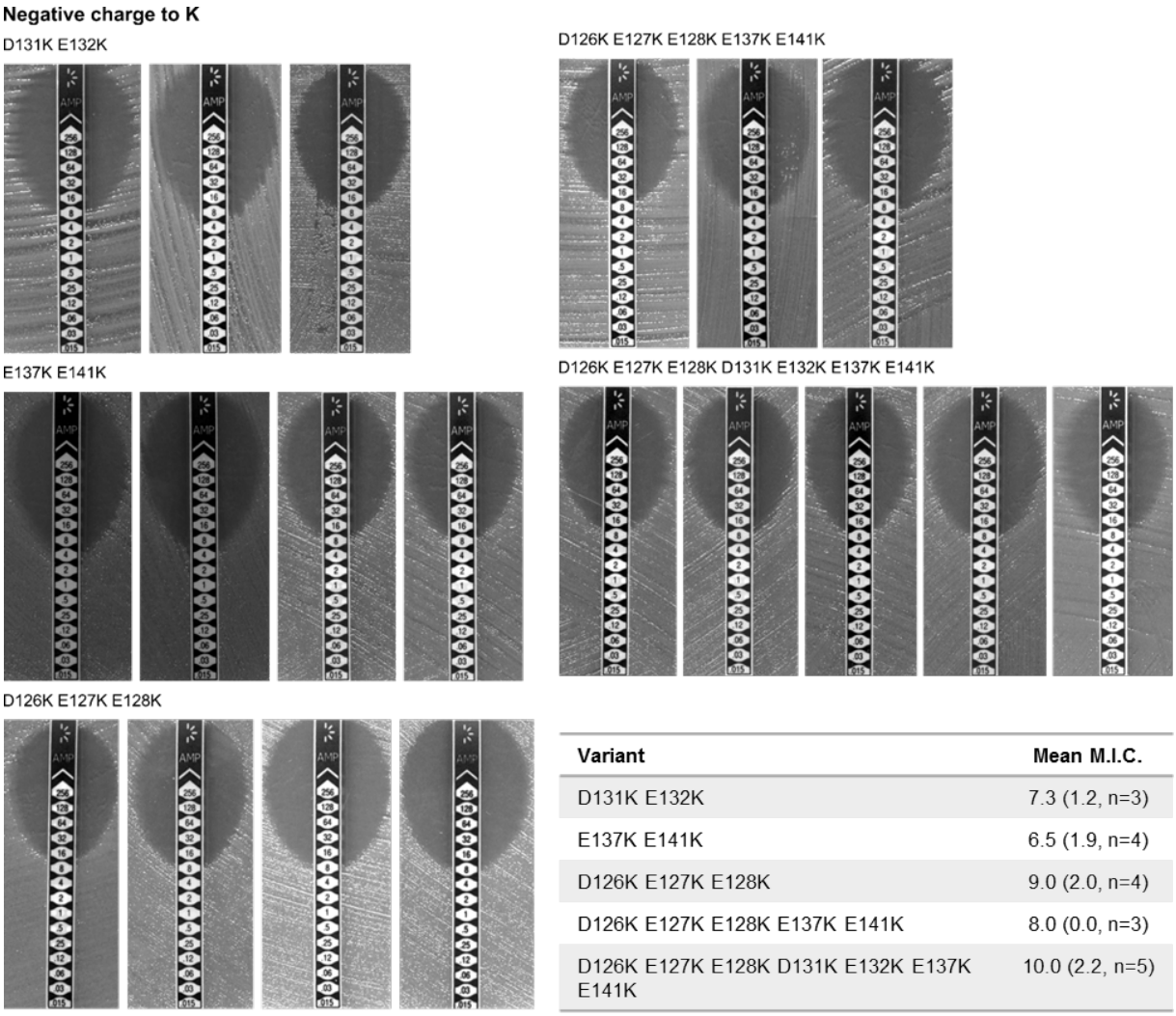


D126A E127A E128A



Variant	Mean M.I.C.
Ins D129 E130 E131	6.7 (1.2, n=3)
D131A E132A	6.0 (2.3, n=6)
E137A E141A	6.7 (2.3, n=5)
D126A E127A E128A	7.2 (1.1, n=4)

Figure 8.11 M.I.C. analysis of negatively charged variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding negatively charged variants (InsD129E130E131-Rieske-Bla, D131AE132A-Rieske-Bla, E137AE141A-Rieske-Bla, and D126AE127AE128A-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

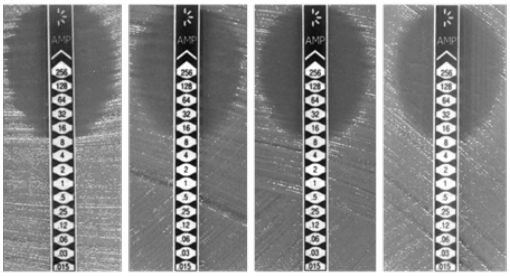


Variant	Mean M.I.C.
D131K E132K	7.3 (1.2, n=3)
E137K E141K	6.5 (1.9, n=4)
D126K E127K E128K	9.0 (2.0, n=4)
D126K E127K E128K E137K E141K	8.0 (0.0, n=3)
D126K E127K E128K D131K E132K E137K E141K	10.0 (2.2, n=5)

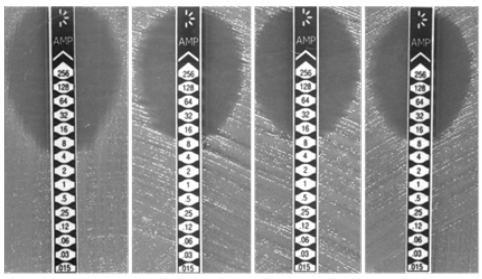
Figure 8.12 M.I.C. analysis of negatively charged variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding negatively charged variants (D131KE132K-Rieske-Bla, E137KE141K-Rieske-Bla, D126KE127KE128K-Rieske-Bla, D126KE127KE128KE137KE141K-Rieske-Bla and D126KE127KE128KD131KE132KE137KE141K - Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The table indicates the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

Truncations - 5aa

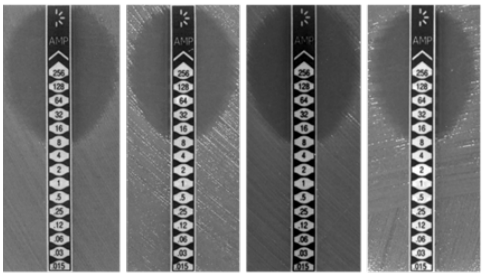
Δ118-122



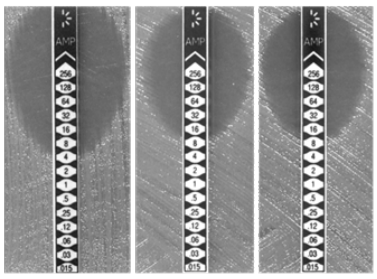
Δ138-142



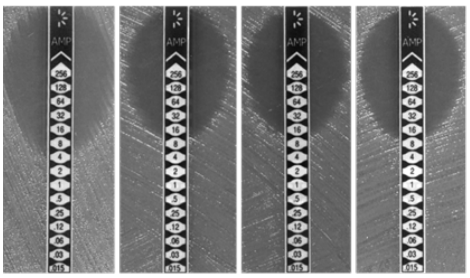
Δ123-127



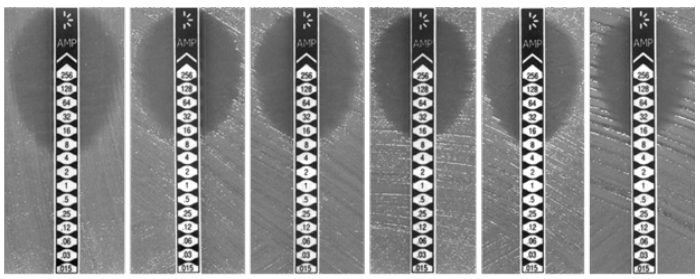
Δ143-147



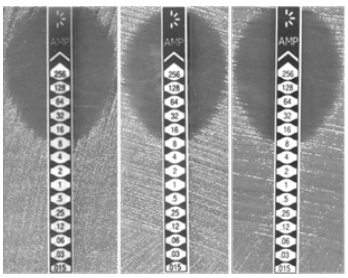
Δ128-132



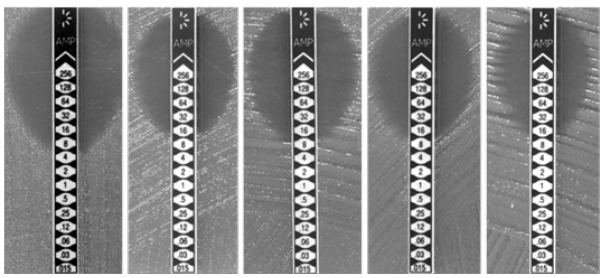
Δ148-152



Δ133-137



Δ153-157



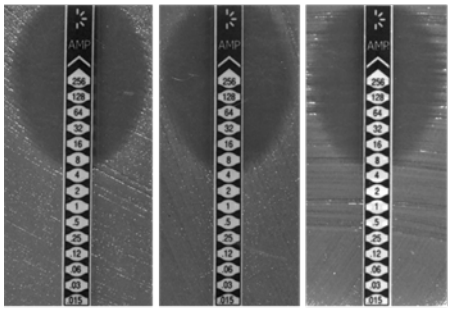
Variant	Mean M.I.C.
Δ118-122	10.0 (2.3, n=4)
Δ123-127	5.3 (2.3, n=4)
Δ128-132	12.0 (0.0, n=4)
Δ133-137	6.7 (2.3, n=3)

Variant	Mean M.I.C.
Δ138-142	7.0 (2.0, n=4)
Δ143-147	6.7 (2.3, n=3)
Δ148-152	10.0 (3.3, n=6)
Δ153-157	9.6 (3.6, n=5)

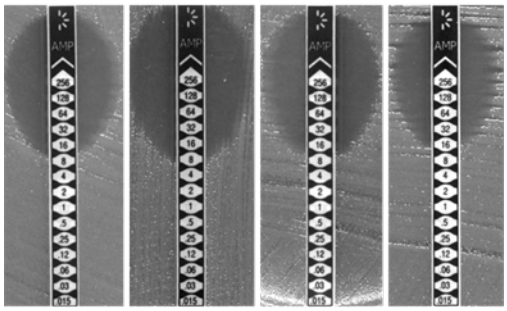
Figure 8.13 M.I.C. analysis of 5aa truncation variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding truncations of 5aa variants (Δ118-122-Rieske-Bla, Δ123-127-Rieske-Bla, Δ128-132-Rieske-Bla, Δ133-137-Rieske-Bla, Δ138-142-Rieske-Bla, Δ143-147-Rieske-Bla, Δ148-152-Rieske-Bla and Δ153-157-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (μg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.

Truncations - 10aa

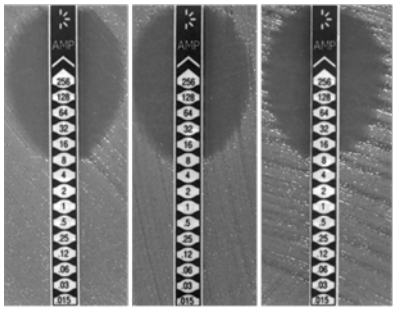
Δ118-127



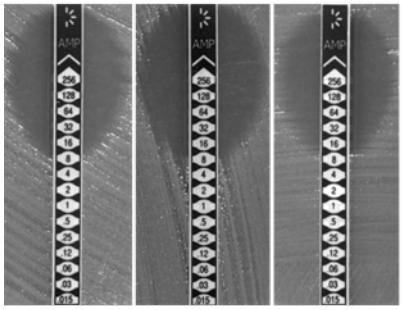
Δ128-137



Δ138-147

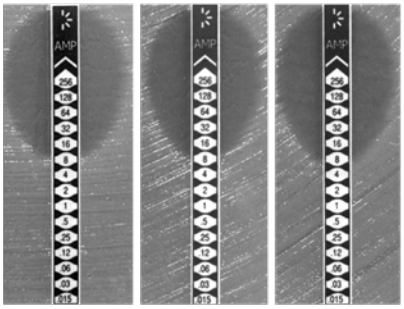


Δ148-157

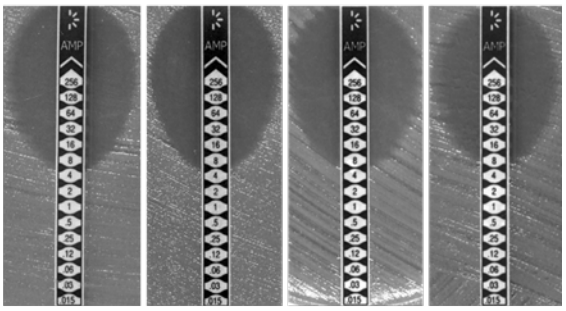


Truncations - 20aa

Δ118-137

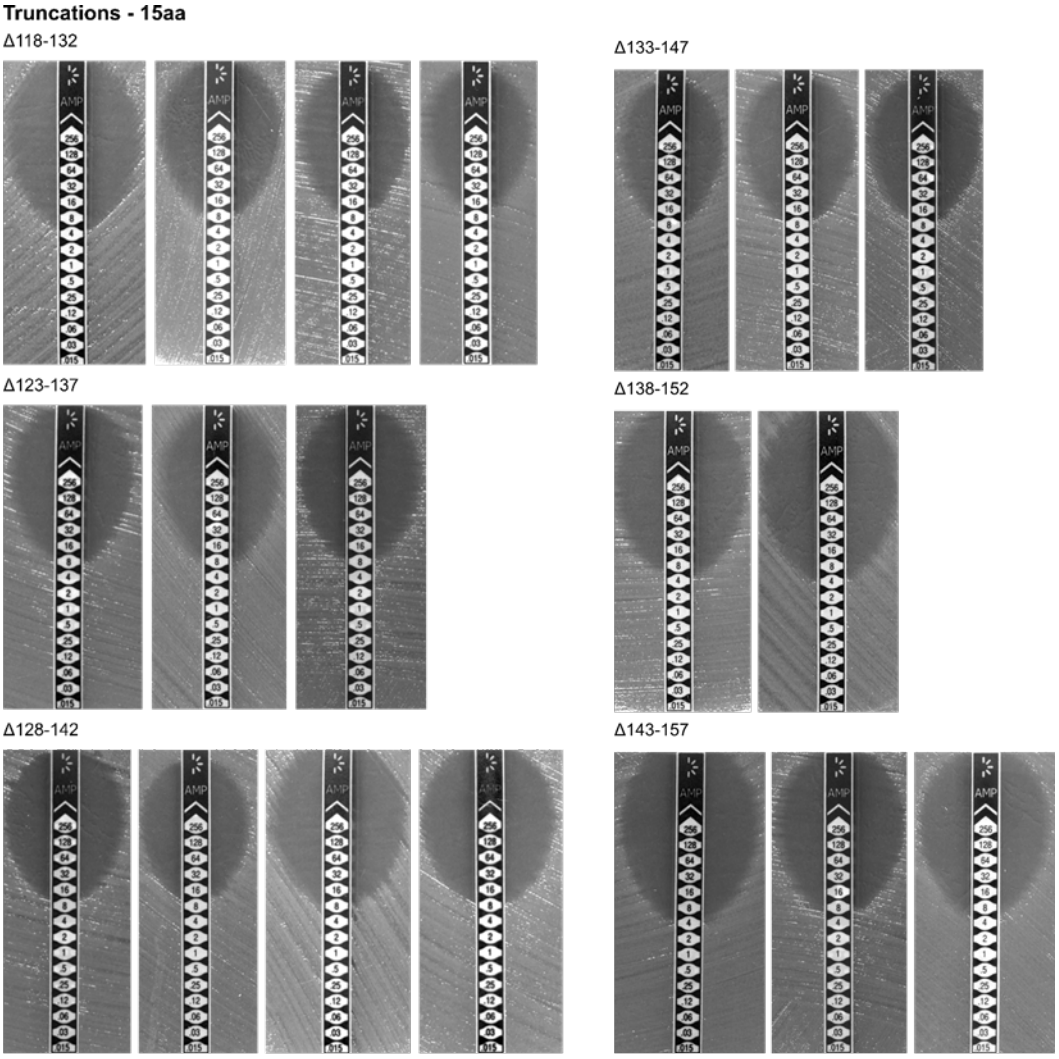


Δ138-157



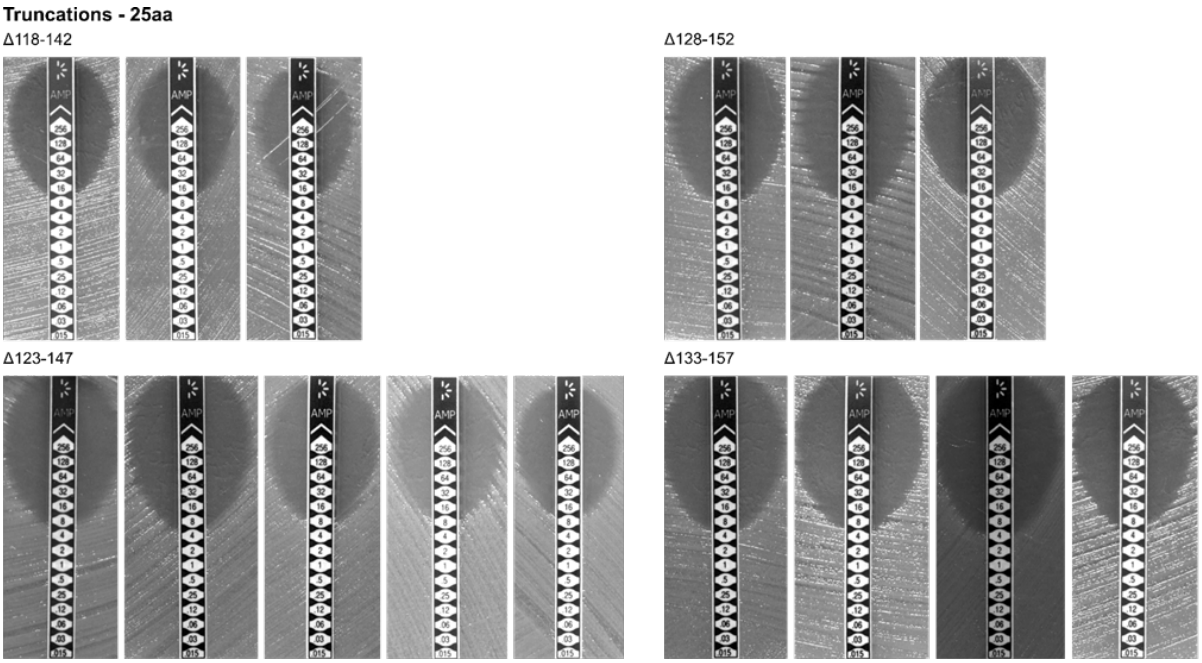
Variant	Mean M.I.C.
Δ118-127	6.0 (0.0, n=3)
Δ128-137	9.5 (3.0, n=4)
Δ138-147	8.0 (0.0, n=3)
Δ148-157	9.3 (2.3, n=3)
Δ118-137	7.3 (1.2, n=3)
Δ138-157	5.5 (1.9, n=4)

Figure 8.14 M.I.C. analysis of 10aa and 20aa truncation variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding truncations of 10aa (Δ118-127-Rieske-Bla, Δ123-137-Rieske-Bla, Δ138-147-Rieske-Bla, Δ148-157-Rieske-Bla) or 20aa variants (Δ118-137-Rieske-Bla and Δ138-157-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.



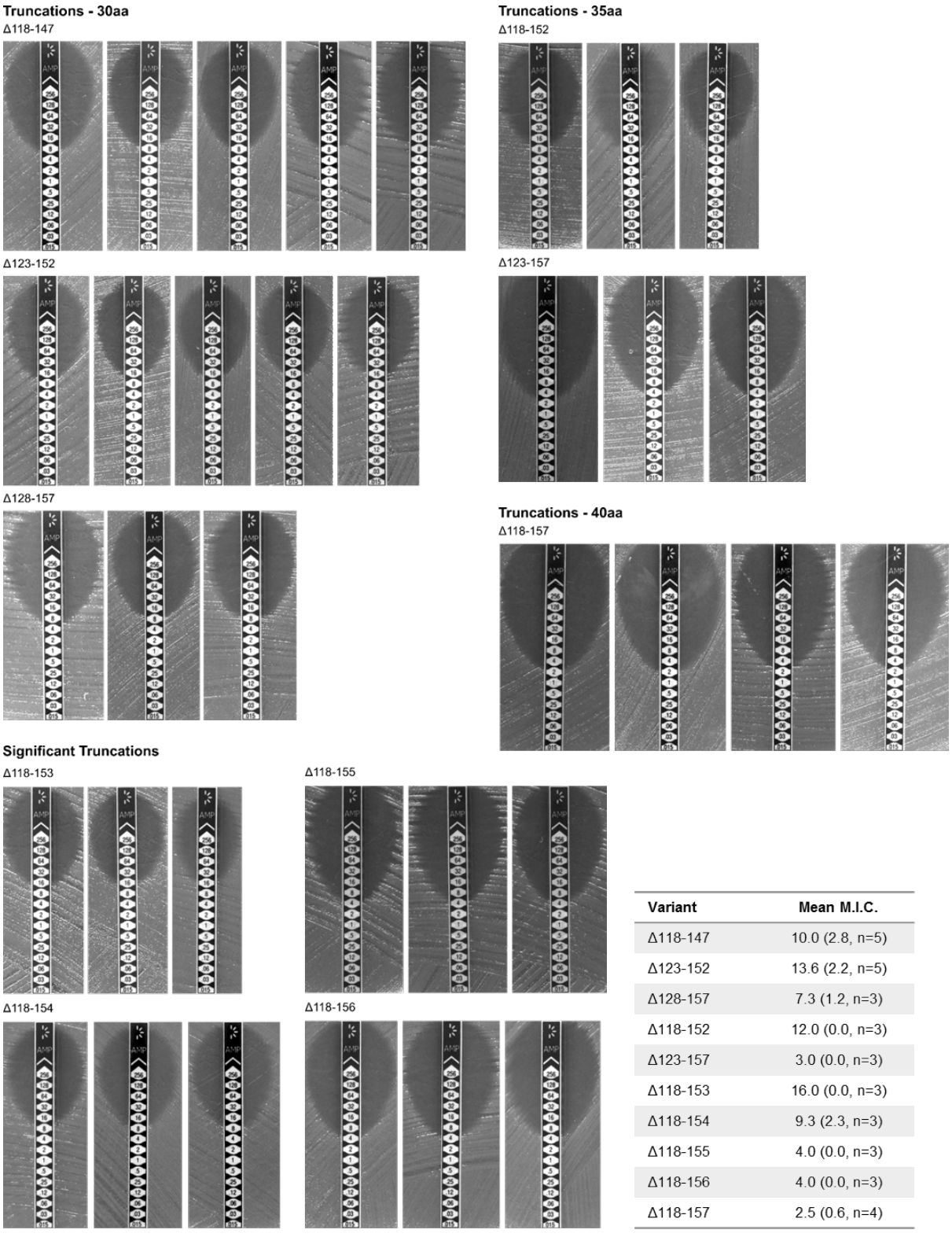
Variant	Mean M.I.C.
Δ118-132	9.0 (3.8, n=4)
Δ123-137	8.0 (0.0, n=3)
Δ128-142	10.0 (2.3, n=4)
Δ133-147	9.3 (2.3, n=3)
Δ138-152	5.0 (1.4, n=2)
Δ143-157	4.8 (1.5, n=4)

Figure 8.15 M.I.C. analysis of 15aa truncation variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding truncations of 15aa variants (Δ118-132-Rieske-Bla, Δ123-137-Rieske-Bla, Δ128-142-Rieske-Bla, Δ133-147-Rieske-Bla, Δ138-152-Rieske-Bla, and Δ143-157-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicates the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.



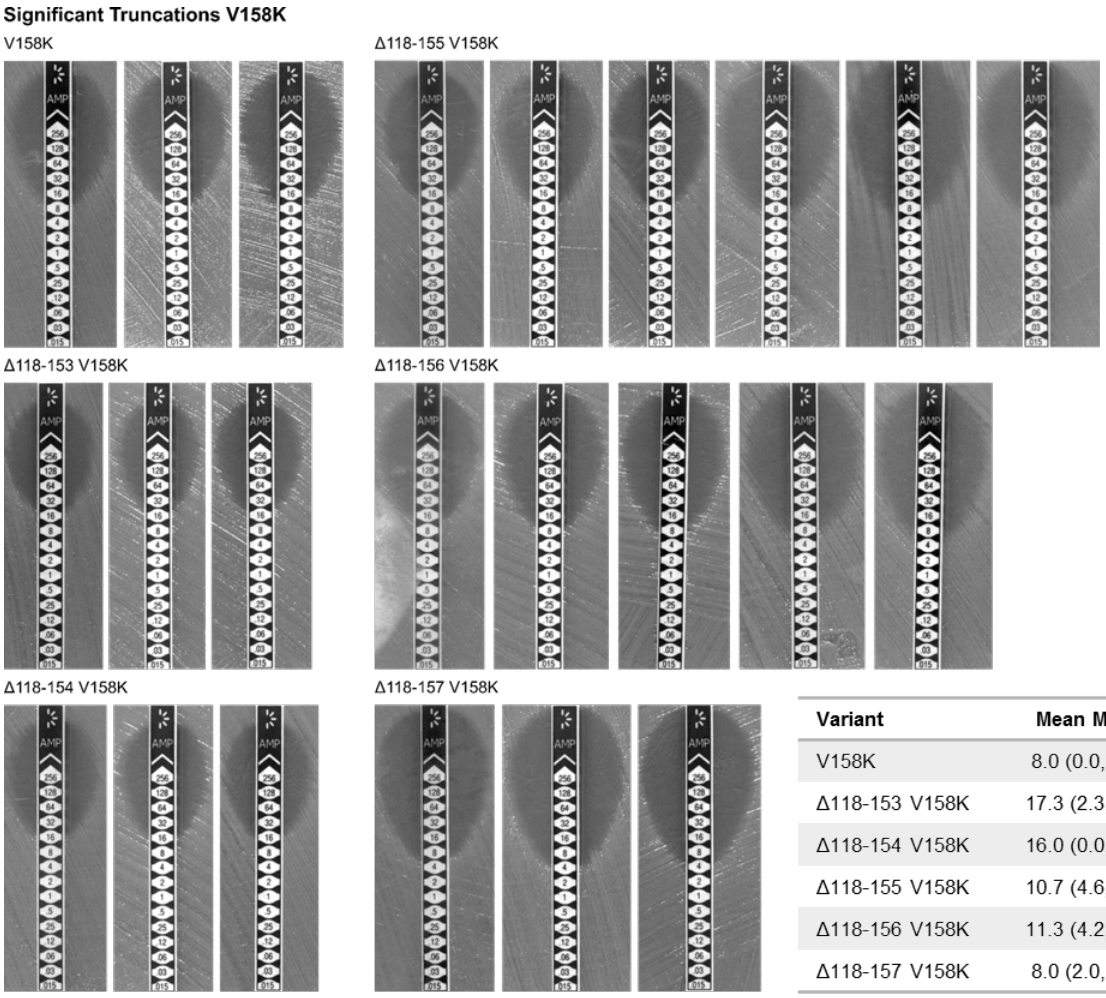
Variant	Mean M.I.C.
Δ118-142	12.0 (0.0, n=3)
Δ123-147	7.6 (0.9, n=5)
Δ128-152	8.0 (0.0, n=3)
Δ133-157	6.0 (0.0, n=4)

Figure 8.16 M.I.C. analysis of 25aa truncation variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding truncations of 25aa variants (Δ118-142-Rieske-Bla, Δ123-147-Rieske-Bla, Δ128-152-Rieske-Bla, and Δ133-157-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The table indicates the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.



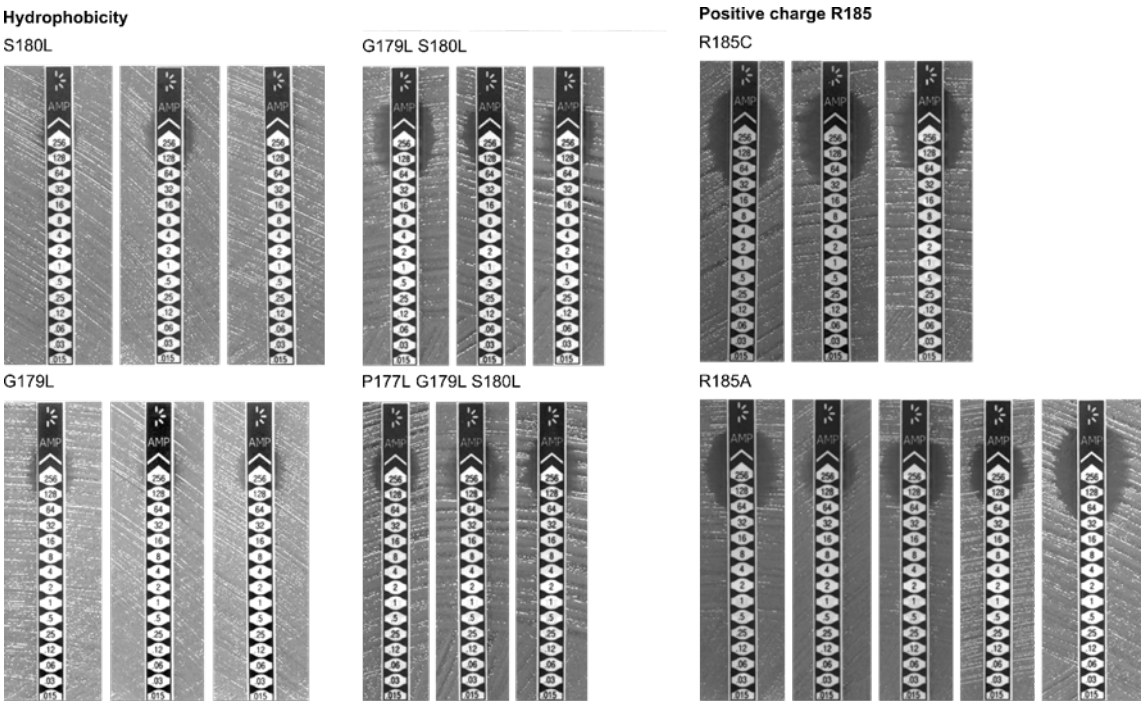
Variant	Mean M.I.C.
Δ118-147	10.0 (2.8, n=5)
Δ123-152	13.6 (2.2, n=5)
Δ128-157	7.3 (1.2, n=3)
Δ118-152	12.0 (0.0, n=3)
Δ123-157	3.0 (0.0, n=3)
Δ118-153	16.0 (0.0, n=3)
Δ118-154	9.3 (2.3, n=3)
Δ118-155	4.0 (0.0, n=3)
Δ118-156	4.0 (0.0, n=3)
Δ118-157	2.5 (0.6, n=4)

Figure 8.17 M.I.C. analysis of 30aa to 40aa truncation variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding truncations of 30aa (Δ118-147-Rieske-Bla, Δ123-152-Rieske-Bla, Δ128-157-Rieske-Bla) or 35aa (Δ118-152-Rieske-Bla and Δ123-157-Rieske-Bla) or significant truncation variants (Δ118-153-Rieske-Bla, Δ118-154-Rieske-Bla, Δ118-155-Rieske-Bla, Δ118-156-Rieske-Bla and Δ118-157-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicates the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.



Variant	Mean M.I.C.
V158K	8.0 (0.0, n=3)
Δ118-153 V158K	17.3 (2.3, n=3)
Δ118-154 V158K	16.0 (0.0, n=3)
Δ118-155 V158K	10.7 (4.6, n=6)
Δ118-156 V158K	11.3 (4.2, n=5)
Δ118-157 V158K	8.0 (2.0, n=3)

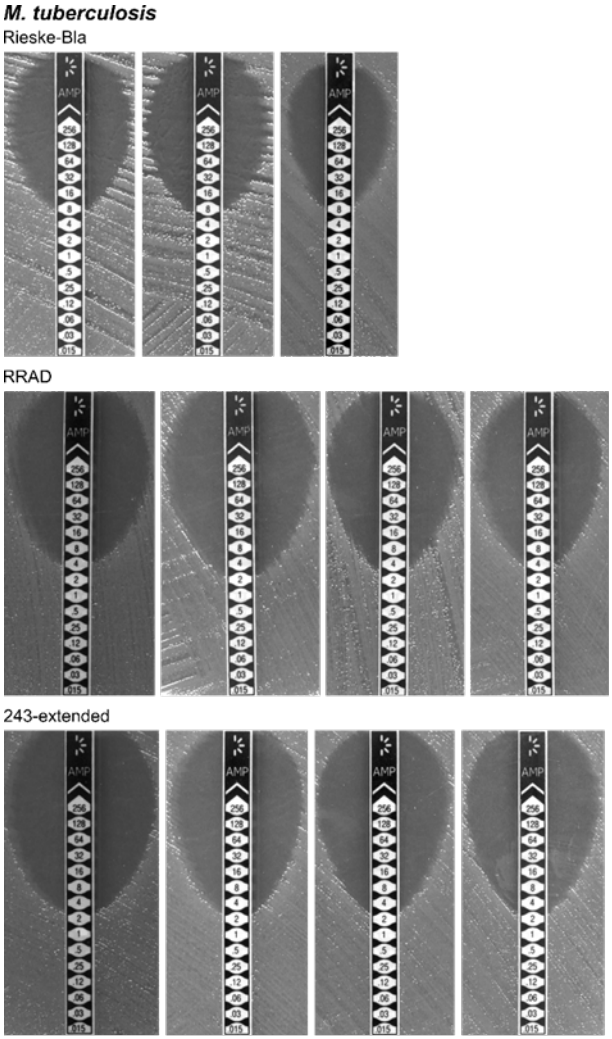
Figure 8.18 M.I.C. analysis of truncation variants with V158K.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding significant truncation variants (V158K, Δ118-153 V158K-Rieske-Bla, Δ118-154V158K-Rieske-Bla, Δ118-155V158K-Rieske-Bla, Δ118-156V158K-Rieske-Bla and Δ118-157V158K-Rieske-Bla). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.



Variant	Mean M.I.C.
S180L	192.0 (64.0, n=3)
G179L	149.0 (37.0, n=3)
G179LS180L	94.7 (32.1, n=3)
P177LG179LS180L	116.0 (20.8, n=3)

Variant	Mean M.I.C.
R185C	42.7 (18.5, n=3)
R185A	76.8 (36.5, n=5)

Figure 8.19 M.I.C. analysis of hydrophobicity variants and R185 variants.
M.I.C.Evaluator™ strip test of DADE (*tat*⁻) harbouring the pSUPROM vector encoding hydrophobicity (S180L-Rieske-Bla, G179L-Rieske-Bla, G179LS180L-Rieske-Bla and P177LG179LS180L-Rieske-Bla) or R185 variants (R185C-Rieske-Bla, and R185A-Rieske-Bla,). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The tables indicate the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.



Variant	Mean M.I.C.
Rieske-Bla	8.0 (0.0, n=3)
RRAD	3.0 (1.0, n=4)
243-extended	2.5 (0.5, n=4)

Figure 8.20 M.I.C. analysis of *M. tuberculosis* Rieske-Bla variants.
M.I.C.Evaluator™ strip test of DADE (*tat*) harbouring the pSUPROM vector encoding *M. tuberculosis* variants (Rieske-Bla, RRAD-Rieske-Bla and 243-extended-Rieske-Bla,). Stationary phase cultures were diluted to OD_{600} 0.1 and a lawn of bacteria was spread onto LB agar plates, M.I.C.Evaluator™ strips were placed on the lawn and the plate was incubated at 37°C for 18 h. The M.I.C. (µg/ml) for ampicillin is read at the intersection of the test strip and the clearing of bacteria. The table indicates the mean M.I.C. and in brackets the s.d. and number of biological replicates, n.